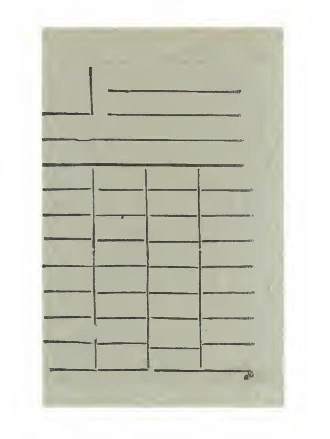


Г. М. Островский,
Т. А. Бережинский

Оптимизация химико-технологических процессов

Теория и практика



x

Моск

Химическая Кибернетика

Г. М. Островский,
Т. А. Бережинский

Оптимизация химико-технологических процессов

Теория и практика



Москва Химия 1984

6П7.1 : 6Ф01
О77

УДК 66.011.001 (024)

Островский Г. М., Бережинский Т. А.

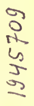
Оптимизация химико-технологических процессов. Теория и практика. — М.: Химия, 1984. — (Серия «Химическая кибернетика»), 240 с., ил.

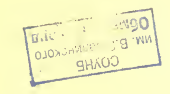
Книга посвящена проблеме оптимизации химико-технологических процессов, возникающей при проектировании новых процессов и интенсификации действующих производств, а также при разработке автоматизированных систем управления технологическими процессами (АСУ ТП). В ней рассматриваются основные этапы этой задачи (расчет стационарных режимов химико-технологических систем, методы безусловной минимизации, алгоритмы учета ограничений), приводятся многочисленные примеры использования описанных методов при решении тестовых и реальных задач оптимизации химико-технологических процессов. Большое внимание уделено проблеме синтеза химико-технологических систем — новому и быстро развивающемуся разделу теории математического моделирования.

Книга предназначена для научных и инженерно-технических работников предприятий и организаций химической и смежных с ней отраслей промышленности. Она представляет интерес для преподавателей, аспирантов и студентов вузов, специализирующихся в области математического моделирования, автоматизации и автоматизированного проектирования химико-технологических процессов.

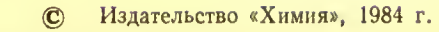
240 с.; 33 табл.; 48 рис.; 166 литературных ссылок.

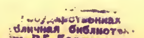
Рецензент — докт. техн. наук Цирлин А. М.





О $\frac{2801020000-059}{050(01)-84}$ 59.84





СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

В книге описываются современные методы оптимизации отдельных аппаратов и химико-технологических систем (ХТС). В ней рассмотрены два класса оптимизационных задач химической технологии: к первому классу относятся задачи оптимизации ХТС фиксированной структуры, ко второму — задачи выбора оптимальной структуры ХТС (синтез ХТС). Эти задачи возникают как при интенсификации действующих, так и при создании новых химико-технологических процессов, в том числе при разработке автоматизированных систем управления технологическими процессами (АСУ ТП). Несмотря на то, что методы решения задач синтеза ХТС начали развиваться в самое последнее время, их разработка стала одной их важнейших проблем математического моделирования химико-технологических процессов. Решение задач обоих классов должно стать неотъемлемой частью создания высокоэффективных химико-технологических процессов.

Рост сложности и размерности этих задач, особенно задач 2-го класса, требует применения наиболее эффективных (как по быстродействию, так и по надежности определения наилучшего решения) методов оптимизации, позволяющих решать эти задачи в реальное время. Гибкость и универсальность поисковых методов оптимизации, относящихся к классу численных методов нелинейного программирования, сделали их основным средством решения задач 1-го класса и существенной частью алгоритмов решения задач 2-го класса. В последнее время такие методы получили большое развитие, особенно это относится к квазиньютоновским методам, и к методам оптимизации больших систем. Основное внимание в книге уделяется этим методам и опыту их использования для оптимизации ХТС. Вместе с тем комбинаторная природа задач синтеза ХТС требует применения методов дискретной математики, использованию которых также уделено большое внимание.

Авторы выражают благодарность сотрудникам НИФХИ им. Л. Я. Карпова и других организаций, оказавшим помощь при подготовке следующих разделов: «Методы сопряженных направлений» (А. Р. Беляевой), «Расчет стационарных режимов химико-технологической схемы изомеризации *н*-пентана» (Н. Н. Зиятдинову и В. Б. Покровскому), «Оптимизация процесса полимеризации изопрена в производстве синтетического каучука» (С. Л. Подвальному и Е. М. Михайловой), «Расчет отделения синтеза аммиака» (Д. Н. Мотылю), «Оптимизация конструкционных параметров в теплообменной системе» (Г. В. Михайлову и В. С. Виткову).

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

E^n — n-мерное евклидово пространство

x_i; p_i; ... — i-тый вектор (вектор-столбец) x; p; ... в последовательности $\{x_i\}$; $\{p_i\}$; ... или i-тая компонента одиночного вектора x; p; ...

H_i; K_i; ... — i-тая матрица H; K; ...

$x_{i,k}$; $y_{i,k}$; ... — i-тая компонента (координата) k-го вектора x; y; ...

h_{ij} — элемент i-той строки и j-го столбца матрицы $H = (h_{ij})$

$x^{\text{т}}$; $H^{\text{т}}$; ... — транспонированный вектор x (вектор-строка); транспонированная матрица $H : H = (h_{ij}) \Rightarrow H^{\text{т}} = (h_{ji})$

n-вектор — вектор из E^n

$(n \times m)$-матрица — матрица из n строк и m столбцов

$\frac{\partial \varphi}{\partial x}$, $\varphi \in E^m$, $x \in E^n$ — $(m \times n)$ = матрица $\left(\frac{\partial \varphi_i}{\partial x_j}\right)$

$\| A \|_F$ — норма Фробениуса (евклидова норма) матрицы $A : \| A \|_F = \sum_{i,j} a_{ij}^2 = \text{Tr}\,(A^{\text{т}} A)$

I; I_n — единичная матрица; единичная $(n \times n)$-матрица

$\overline{1, n}$ — 1, 2, ..., n

$\forall \alpha$ — для всех α

$\exists x$ — существует x

K_f — число обращений к расчету функции f

K_p — число векторов p

δ_{ij} — символ Кронекера: $\delta_{ij} = 1$, если $i = j$, $\delta_{ij} = 0$, если $i \neq j$

ГЛАВА I

Постановка задачи оптимизации сложных химико-технологических систем

Современное химико-технологическое производство представляет собой систему взаимосвязанных аппаратов. Оптимизация отдельных аппаратов без учета их связей с остальными аппаратами может привести к неоптимальной работе всей химико-технологической системы (ХТС). Отсюда возникает задача оптимизации всей системы в целом, в которой учитывается взаимное влияние аппаратов. Как и при постановке задачи оптимизации в любой другой области, здесь необходимо сформулировать математическую модель системы, критерий оптимизации и ограничения на переменные. Математическая модель ХТС состоит из двух частей — совокупности математических моделей отдельных блоков и математической модели структуры ХТС. Математическая модель отдельного блока имеет вид

$$z^{(k)} = f^{(k)}\left(x^{(k)},\ u^{(k)}\right),\ k = \overline{1, N} \tag{I,1}$$

где $z^{(k)} = \left(z_1^{(k)},\ \ldots,\ z_{m_k}^{(k)}\right)$, $x^{(k)} = \left(x_1^{(k)},\ \ldots,\ x_{n_k}^{(k)}\right)$, $u^{(k)} = \left(u_1^{(k)}, \ldots, u_{r_k}^{(k)}\right)$ — m_k-вектор выходных переменных; n_k — вектор входных переменных, r_k — вектор управляющих переменных. Переменные $z^{(k)}$, $x^{(k)}$ характеризуют состояние выходных и входных потоков k-го блока, а в число переменных $u^{(k)}$ включаются конструктивные и технологические параметры, изменяя которые можно влиять на режим работы k-го блока. Как правило, в число переменных $x^{(k)}$, $z^{(k)}$ включаются концентрации компонент потока, температура, давление и расход потока. Величины $x^{(k)}$, $z^{(k)}$ часто называют переменными состояния, или фазовыми переменными. Здесь, естественно, не предполагается известной явная аналитическая зависимость выходных переменных от входных и управляющих переменных. Исходят только из того, что имеется алгоритм, позволяющий по известным входным переменным и управлениям подсчитать выходные переменные k-го блока. Так, для определения выходных переменных реактора с неподвижным слоем придется решать задачу Коши для системы обыкновенных дифференциальных уравнений (а в случае учета продольной диффузии — даже краевую задачу) [1, с. 33], для определения выходных переменных реактора идеального смешения или любого аппарата с противотоком (ректификационная колонна, абсорбер и т. д.) — систему нелинейных конечных уравнений и т. д.

Рассмотрим теперь математическую модель структуры ХТС. Отметим вначале, что потоки ХТС можно разделить на три группы — входные, промежуточные и выходные потоки. Входным будем считать поток, который подается извне в один из блоков ХТС, выходным — тот, который выходит из блока ХТС и подается куда-либо вне схемы, и, наконец, промежуточным — поток, который выходит из одного блока ХТС и подается в другой ее блок. Переменные, характеризующие входные, выходные и промежуточные потоки ХТС, будем называть входными, выходными и промежуточными переменными схемы, соответственно. Далее предположим, что в каждом блоке первые s_k ($s_k \leqslant n_k$) входных переменных являются входными переменными системы, и первые g_k ($g_k \leqslant m_k$) выходных переменных — выходными переменными системы. Конечно, в большинстве блоков $s_k = g_k = 0$. Вектор, компонентами которого являются входные переменные блока, являющиеся входными переменными системы, обозначим через $x^{(k)} = \left(x_1^{(k)}, \ \ldots, \ x_{s_k}^{(k)}\right)$. Для описания структуры ХТС пользуются моделями двух типов. В первом случае структура ХТС описывается системой соотношений вида

$$x_i^{(k)} - z_{g_{k,i}}^{h_{k,i}} = 0 \quad i = s_k + 1, \ldots, n_k \quad k = \overline{1,N} \qquad (I,2)$$

Такая запись означает, что $g_{k,i}$-я выходная переменная $h_{k,i}$-го блока равна i-й входной переменной k-го блока. Эта запись предполагает, что в разных потоках число компонент может быть различно. Соотношения (I,2) характеризуют связи всех входных и выходных промежуточных переменных блоков. Рассмотрим математические модели структуры ХТС другого вида. Пусть k-й блок имеет N_k входных и M_k выходных потоков. Обозначим через $x^{(ks)} = \left(x_1^{(ks)}, \ \ldots, \ x_{n_{hs}}^{(ks)}\right)$ вектор переменных s-го входного потока k-го блока и через $z^{(kj)} = \left(z_1^{(kj)}, \ \ldots, \ z_{m_{kj}}^{(kj)}\right)$ — вектор переменных j-го выходного потока k-го блока. Математическая модель k-го блока может быть теперь записана в виде

$$z^{(kj)} = f^{(kj)}\left(x^{(k1)}, \ldots, x^{(kN_k)}\right) \quad j = \overline{1,M_k} \qquad (I,3)$$

Далее удобно будет считать, что все потоки, т. е. все векторы $x^{(ki)}$, $z^{(kj)}$ обладают одними и теми же компонентами. Это, конечно, не является дополнительным допущением, поскольку, если в некотором потоке имеются не все компоненты, то концентрации отсутствующих компонент формально можно считать равными нулю. Отсюда соотношения связи между блоками могут быть записаны в виде

$$x^{(ks)} = \sum_{i=1}^{N} \sum_{j=1}^{N_k} \alpha_{ki}^{sj} z^{(ij)} \qquad (I,4)$$

Параметры α_{ki}^{sj} называются структурными и определяются из условия:

$$\alpha_{ki}^{sj} = \begin{cases} 1, & \text{если } j\text{-тый выходной поток } i\text{-го блока} \\ & \text{подается на } s\text{-й вход } k\text{-го блока} \\ 0, & \text{в противном случае} \end{cases} \qquad (1,4\text{a})$$

Если i-тый блок будет иметь один выходной поток, а k-тый блок — один входной поток, пользуются более простым обозначением для структурных параметров:

$$\alpha_{ki}^{11} = \alpha_{ki} \tag{I,5}$$

В том случае, если каждый блок имеет один входной и выходной потоки, соотношения (I, 4) с учетом обозначений (I, 5) примут вид

$$x^{(k)} = \sum_{j=1}^{N} \alpha_{kj} z^{(j)} \tag{I,6}$$

Выходной поток любого блока может подаваться только на один вход другого блока. Отсюда, если, например $\alpha_{lm}^{pq} = 1$, то $\alpha_{km}^{sq} = 0$ для всех $k \neq l$, $s \neq p$. Это условие можно выразить соотношением

$$\psi^{q,m} = \sum_{s,k} \alpha_{km}^{sq} - 1 = 0 \quad m = \overline{1,N} \quad q = \overline{1,M_k} \tag{I,7}$$

На управления в каждом блоке налагаются ограничения вида

$$\varphi_i^k \left(u^k\right) \leqslant a^{(k)} \quad i = \overline{1,p_k} \tag{I,8}$$

В большинстве случаев они имеют следующий простой вид

$$c^{(k)} \leqslant u^{(k)} \leqslant b^{(k)} \tag{I,9}$$

где $c^{(k)}$, $b^{(k)}$ — некоторые заданные r_k-векторы. Обычно налагаются также ограничения на выходные переменные системы. Для простоты записи будем считать, что первые $\bar{g}_k$ $(\bar{g}_k < g_k)$ выходных переменных блока фиксированы

$$z_i^{(k)} = a_i^{(k)} \quad i = \overline{1,\bar{g}_k} \quad k = \overline{1,N} \tag{I,10}$$

Более сложные ограничения на выходные переменные легко могут быть сведены к ограничениям типа (I, 10).

Для простоты изложения переменные $z_i^{(k)}$ $(i = \overline{1, \bar{g}_k})$ будем иногда называть закрепленными выходными переменными k-го блока, а переменные $z_i^{(k)}$ $(k = \bar{g}_k + 1, \ldots, g_k)$ — свободными выходными переменными k-го блока. Иногда имеются также ограничения на промежуточные переменные

$$\varphi\left(x^{(k)}\right) \leqslant 0 \tag{I,11}$$

$$\psi\left(z^{(k)}\right) \leqslant 0 \tag{I,12}$$

В дальнейшем, если не оговорено противное, будем предполагать, что все входные переменные схемы $x_i^{(k)}$ $(i = \overline{1, s_k} \quad k = \overline{1, N})$ являются свободными переменными, подлежащими определению при решении задачи оптимизации ХТС. Эти свободные входные переменные формально ничем не отличаются от управлений, поэтому те и другие иногда будем считать управлениями, вводя при этом обозначения:

$$x_i^{(k)} = u_{r_k+i}^{(k)} \quad i = \overline{1,s_k} \quad k = \overline{1,N} \tag{I,13}$$

Пусть

$$\bar{r}_k = r_k + s_k \tag{I,14}$$

Тогда вектор $u^{(k)}$ будет $\bar{r}_k$-вектором. Критерий оптимизации ХТС обычно является аддитивной функцией, зависящей от свободных входных и выходных переменных состояния, а также управлений. В общем случае он имеет вид:

$$F = \sum_{k=1}^{N} F^{(k)} \left(u^{(k)}, \bar{\bar{z}}^{(k)}\right) \tag{I,15}$$

где $\bar{\bar{z}}^{(k)} = \left(z_{\bar{g}_k+1}, \ \ldots, \ z_{g_k}\right)$ — вектор свободных выходных переменных k-го блока; $F^{(k)}$ — часть критерия, относящаяся к k-му блоку. Задача оптимизации ХТС ставится следующим образом. Необходимо определить управления $u^{(k)}$ и переменные состояния $z^{(k)}$, $x^{(k)}$ таким образом, чтобы при выполнении уравнений (I,1), (I,2) или (I,3), (I,4), а также ограничений (I,8), (I,10) критерий (I,15) принял минимальное значение.

Задача опимизации процесса разделения изомеров октана

Химико-технологический процесс [2], схема которого дана на рис. 1, предназначен для выделения компонент A_1 и A_3 из смеси изомеров октана A_1—A_4. На вход узла разделения подается указанная смесь изомеров вместе с небольшим количеством углеводородных смесей C_{7-} и C_{9+}. Входной поток распределяется между блоком разделения и кристаллизатором. Некоторая его часть может быть направлена прямо на выход из системы (к бензосмесительной станции). Двухступенчатая схема разделения, включающая две ректификационные колонны K-1 и K-2, предназначена для отделения изомера A_3. Тяжелые ароматические соединения, собирающиеся в кубе второй колонны K-2, также являются одним из видов товарной продукции. Дистиллят, отбираемый из первой колонны K-1 и содержащий в основном смесь изомеров октана A_1, A_2, A_4 и углеводородов C_{7-}, C_{9+}, распределяется между кристаллизатором и блоком изомеризации. Некоторая его часть может быть также выведена из системы на бензосмесительную станцию. Определенная часть дистиллята смешивается с циркулирующим потоком, возвращаемым из блока изо-

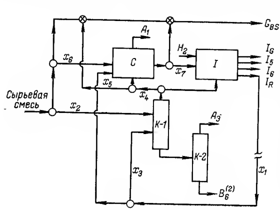


Рис. 1. Схема процесса разделения изомеров октана:
К-1, К-2 — ректификационые колонны; C — кристаллизатор; I — блок изомеризации; G_{BS} — поток на бензосмесительную станцию; I_G, I_5, I_6, I_R — потоки соответственно остаточного газа, легких и тяжелых углеводородов, циркулирующей смеси; $B_6^{(2)}$ — поток тяжелых углеводородов.

меризации и содержащим A_1—A_4, и некоторой долей входного потока и подается на вход кристаллизатора. Из верхней части кристаллизатора отбирается изомер A_1, практически не содержащий примесей других углеводородов, а остальная часть нефтепродуктов (рафинат), выходящих из кристаллизатора, распределяется между блоком изомеризации и бензосмесительной станцией, в которую, как было отмечено ранее, может направляться также некоторая доля потока, входящего в систему, и часть дистиллята из первой колонны K-1. Доля потока из кристаллизатора, направляемая в блок изомеризации, на входе в аппарат смешивается с некоторой долей дистиллята из первой ректификационной колонны K-1.

Для проведения процесса изомеризации в аппарат вводится небольшое количество водорода, составляющее 0,5 % общей нагрузки аппарата. Получаемые на выходе остаточный газ, легкие и тяжелые ароматические соединения представляют собой товарную продукцию, а равновесная смесь изомеров октана A_1—A_4 возвращается в цикл.

Математическая модель процесса *

Математическое описание системы включает модели ректификационных колонн K-1 и K-2, кристаллизатора и блока изомеризации, а также узлов разделения и смешения потоков. Модели аппаратов описываются линейными балансовыми соотношениями. Значения индекса $i = 1$—6 присвоим соответственно компонентам потока A_1—A_4, C_{7-}, C_{9+}.

Ректификационные колонны K-1, K-2. Обозначим через $T_{k,i}$ i-тую компоненту смеси на входе и в дистилляте k-той ($k = 1, 2$) колонны, через $B_{k,i}$ — i-тую компоненту в кубовом остатке k-той колонны ($k = 1, 2$), а через x, y с нижними индексами — соответственно входные и выходные величины потоков компонент смеси. Так, $y_{T_{k,i}}$ означает количество i-той компоненты смеси в дистилляте k-той колонны, а $x_{T_{k,j}}$ — количество j-той компоненты на входе в k-тую колонну. Аппараты описываются следующими уравнениями:

$$y_{T_{k,i}} = a_{T_{k,i}} x_{T_{k,i}} \quad y_{B_{k,i}} = x_{T_{k,i}} - y_{T_{k,i}} \qquad (I,16)$$

где $a_{T_{k,i}}$, $i = 1$—6, $k = 1, 2$ — коэффициенты, значения которых приведены в табл. 1.

Кристаллизатор. Введем обозначения: C_i — i-тая компонента смеси на входе в аппарат и на выходе из верхней его части; R_i — i-тая компонента на выходе из нижней его части; x и y с нижними индексами — входные и выходные величины потоков компонент для данного аппарата; θ — теоретический параметр процесса кристаллизации

$$\theta = 1,0698 x_C - 1,0758 x_{C_1} \qquad (I,17)$$

* Математическая модель и экономические характеристики процесса приведены по зарубежной работе [2].

Таблица 1. **Коэффициенты** $a_{T_{k},\ i}$

k	Коэффициенты для $i = 1-6$					
	1	2	3	4	5	6
1	0,9951	0,9869	0,05	0,9998	1	0
2	0,9978	0,9975	0,99	1	—	0,1299

где x_C — полный входной поток; $x_C = \sum_{i=1}^{6} x_{C_i}$. Тогда уравнения модели примут вид:

$$y_{C_1} = 0{,}8\,(x_{C_1} - 0{,}075\theta) \qquad y_{C_2} = 0{,}005639\,(x_{C_1} - 0{,}075\theta)$$
$$y_{C_i} = 0 \qquad i = 3-6 \qquad y_{R_i} = x_{C_i} - y_{C_i} \qquad i = 1-6 \qquad (I,18)$$

Блок изомеризации. Обозначая, как и прежде, через x и y входные и выходные величины потоков, а через x_{H_2} — подачу водорода в аппарат $x_{H_2} = 0{,}005x_I$, где x_I — полный входной поток, и принимая за I_G, I_i, I_R соответственно остаточный газ, i-тую компоненту смеси и смесь изомеров A_1—A_4, запишем уравнения модели в виде:

$$z = x_I + x_{H_2} - y_{I_5} - x_{I_6} \qquad y_{I_G} = 0{,}025z$$
$$y_{I_5} = 0{,}006z + x_{I_5} \qquad y_{I_6} = 0{,}004z + x_{I_6} \qquad y_{I_R} = 0{,}965z \qquad (I,19)$$

При этом состав смеси в циркулирующем потоке I_R полностью определен:

$$y_{I_{R_1}} = 0{,}2y_{I_R} \qquad y_{I_{R_2}} = 0{,}45y_{I_R} \qquad y_{I_{R_3}} = 0{,}25y_{I_R} \qquad y_{I_{R_4}} = 0{,}1y_{I_R} \qquad (I,20)$$

Формулировка задачи оптимизации процесса

Задача оптимизации заключается в том, чтобы получить максимальную денежную отдачу от капитальных вложений в данное производство за счет перераспределения потоков между аппаратами системы; варьируемыми переменными являются x_i, $i = 1-6$ — коэффициенты разделения потоков между двумя трубопроводами в точках из разветвления. При этом состав смеси, подаваемой в узел разделения в количестве x_F, считается заданным:

$$x_{F_1} = 0{,}18x_F \qquad x_{F_2} = 0{,}45x_F \qquad x_{F_3} = 0{,}23x_F$$
$$x_{F_4} = 0{,}11x_F \qquad x_{F_5} = 0{,}01x_F \qquad x_{F_6} = 0{,}02x_F \qquad (I,21)$$

Приведем выражения для составляющих компонент (затрат), с учетом которых формируется в данном случае выражение для критерия оптимизации.

Капитальные вложения определяют так:
для ректификационных колонн

$$J_F = 7,4x_{T_1}^{0,7} + 2,3x_{T_2}^{0,7}$$

для кристаллизатора (с учетом отчислений)

$$J_C = 75y_C^{0,7} + 4,5y_C \quad (I,22)$$

для блока изомеризации (с учетом стоимости катализатора и отчислений)

$$J_I = 6,4x_I^{0,7} + 0,06x_I + 50 \quad (I,23)$$

на тепло-, водо- и энергоснабжение

$$J_U = 0,00956x_{T_1} + 0,002775x_{T_2} + 0,441y_C + 0,05x_I \quad (I,24)$$

на хранение в резервуарах (расходный бак, хранилище продуктов A_1 и A_3) и трубопроводы

$$J_T = 0,2935\,(x_F + y_C + y_{T_2}) \quad (I,25)$$

где x_F — величина входного потока, $x_{T_k} = \sum_{i=1}^{6} x_{T_{k,\,i}}, y_{T_k} = \sum_{i=1}^{6} y_{T_{k,\,i}}$, $k = 1,\ 2$.

Дополнительные сторонние вложения вычисляют по формуле:

$$J_O = 0,1\,(7,4x_{T_1}^{0,7} + 75y_C^{0,7} + 6,4x_I^{0,7} + J_U + J_T) \quad (I,26)$$

Общая сумма капитальных затрат составит

$$J = J_F + J_C + J_I + J_U + J_T + J_O \quad (I,27)$$

Оборотный капитал равен

$$W = 9,75y_{T_2} + 5,25y_C \quad (I,28)$$

Прибыль представляет собой сумму средств от продажи товара за вычетом затрат на его производство.

Приведем выражения для составляющих этих компонент прибыли.

Сумма средств от продажи (в расчете на 1 сут)

$$S = 65y_{T_2} + 30y_C + (26,7y_G - 8,9y_{G_3}) + 17,9y_{I_6} + 18,8\,(y_{I_6} + y_{B_{2,6}}) + 5,8y_{I_G} \quad (I,29)$$

где y_G, y_{G_i}, $i = 1$—6 — соответственно величины общего потока и потока i-той компоненты смеси, направляемого на бензосмесительную станцию.

Стоимость сырья

$$C_C = 24,7x_F + 46,9x_{H_2} \quad (I,30)$$

Заводские затраты (расходы на зарплату, топливо, воду, тепло- и энергоснабжение, химические препараты, катализатор)

$$C_M = 816 + 0,028x_{T_1} + 0,0081x_{T_2} + 3,2y_C + 1,15x_T \quad (I,31)$$

Прочие затраты

$$C_J = 0,15J \quad (I,32)$$

Общая сумма затрат

$$C = C_C + C_M + C_J \qquad (I,33)$$

Критерий оптимизации в расчете на год (%)

$$y = [100 \cdot 365\,(S - C)]/(J + W) \qquad (I,34)$$

Решение задачи оптимизации должно удовлетворять следующим ограничениям:

$$y_C - P = 0 \qquad (I,35)$$

$$110 \leqslant y_{T_{2,3}} \leqslant 274 \qquad (I,36)$$

где P — заданная величина. Результаты решения этой задачи будут приведены в главе IV.

Математический аппарат решения задач оптимизации

Для решения задач оптимизации химико-технологических процессов обычно используют методы нелинейного программирования (поисковые методы) [1, 3] и методы теории оптимального управления: вариационного исчисления [4], динамического программирования [5], принципа максимума Понтрягина [6], дискретного принципа максимума [7]. Наибольшее распространение получили поисковые методы как наиболее гибкие и универсальные. Эти методы находят также широкое применение при решении задач идентификации (определение некоторых коэффициентов уравнений, представляющих собой математическую модель исследуемого процесса). Кроме того, поисковые методы могут быть эффективно использованы при синтезе оптимальной структуры химико-технологических систем, который в общем случае представляет собой задачу дискретно-непрерывного программирования; в частности, они могут быть использованы при получении нижних оценок в методе ветвей и границ (см. гл. VI).

Изложению алгоритмов поисковых методов оптимизации химико-технологических процессов, их теоретических и практических аспектов и посвящена эта книга.

Общая формулировка задачи оптимизации и характеристика методов ее решения

Большое число задач оптимизации химико-технологических процессов может быть представлено в следующем виде:

$$\min_{x \in S} f(x) \qquad (I,37)$$

где f — вещественная функция переменной $x \in E^n$, $S \subseteq E^n$ $(S \neq \varnothing)$. Семейство задач на безусловный минимум, методам решения которых посвящена глава III, соответствует случаю $S \equiv E^n$ в задаче (I, 37). Семейство задач минимизации с ограничениями, представляющее так называемую общую задачу нелинейного программирования, по-

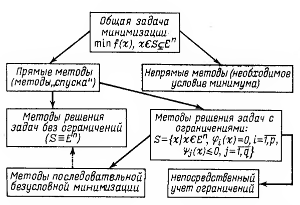


Рис. 2. Классификация методов решения задач минимизации.

лучается из выражения (I,37), если $S \subset E^n$ определяется системой равенств и неравенств

$$S = \{x \mid x \in E^n, \varphi_i(x) = 0, i = \overline{1, p}, \quad \psi_j(x) \leqslant 0, j = \overline{1, q}\} \qquad (I,38)$$

Методы решения задач минимизации можно разделить (в известной степени условно) на две группы (рис. 2). К первой относятся так называемые прямые методы, базирующиеся на непосредственном сравнении значений функции в соседних точках, ко второй — непрямые методы, при использовании которых положение минимума определяется с помощью соответствующего необходимого условия. В дальнейшем всюду речь будет идти лишь о прямых методах решения задач минимизации, т. е. о методах «спуска».

Излагаемые ниже методы носят итерационный характер, т. е. представляют собой совокупность определенных вычислительных процедур с применением рекуррентных формул, результатом выполнения которых является построение конечной или бесконечной последовательности точек $\{x_i\} \in E^n$, $i = 0, 1, \ldots$, позволяющей с заданной точностью найти минимум $f(x)$. В методах безусловной минимизации, т. е. методах решения задач без ограничений, соответствующая последовательность $\{x_i\} \in E^n$ обладает свойством:

$$f(x_0) > f(x_1) > \cdots \geqslant \min f(x)$$

где f — минимизируемая функция.

Большинство методов безусловной минимизации предусматривает построение последовательности $\{p_i\}$, $i = 0, 1, \ldots$ направлений движения к минимуму $f(x)$. Основное отличие одного метода от другого заключается в способе построения p_i, $i = 0, 1, \ldots$ При этом соседние точки последовательности $\{x_i\}$ связаны соотношением (рис. 3):

$$x_{i+1} = x_i + \alpha_i p_i \qquad (I,39)$$

где α_i — параметр, определяющий длину шага вдоль направления p_i, выбираемую некоторым определенным способом.

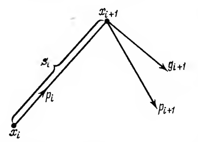


Рис. 3. Графическое изображение итерационного шага.

Вектор «сдвига» s_i в направлении p_i равен

$$s_i = \Delta x_i = x_{i+1} - x_i = \alpha_i p_i \quad (I,40)$$

Таким образом, выполнение одной (i-той) итерации (или шага) в конкретном алгоритме спуска заключается в определении направления спуска p_i и точки x_{i+1} на данном направлении.

Структуру рассматриваемых далее методов безусловной минимизации можно представить следующей схемой (рис. 4).

Шаг 0. Выбирается некоторая начальная точка $x_0 \in E^n$ и положительно определенная $(n \times n)$-матрица H_0. Выполняется расчет градиента g_0 минимизируемой функции в точке x_0: $g_0 = g(x_0)$.

Шаг 1. Даны: точка $x_i \in E^n$, вектор градиента $g_i = g(x_i)$, $(n \times n)$-матрица H_i. Определяется направление

$$p_i = -H_i^{\text{T}} g_i \quad (I,41)$$

В общем случае вектор направления p_i является некоторой явной функцией точки x_i, предыдущих точек x_{i-1}, x_{i-2}, ..., векторов градиентов g_i, g_{i-1}, ... минимизируемой функции f и матриц G_i, G_{i-1} ее вторых производных, вычисленных в точках x_i, x_{i-1}, ..., т. е.

$$p_i = p_i(x_i, x_{i-1}, \ldots; \; g_i, g_{i-1}, \ldots; \; G_i, G_{i-1}, \ldots) \quad (I,42)$$

В выражении (I, 42) зависимость p_i от некоторых групп переменных, например G_i, G_{i-1}, ..., может отсутствовать.

Если существует явная зависимость от предыдущих точек x_{i-1}, ... или величин, вычисленных в этих точках, то подобный поисковый

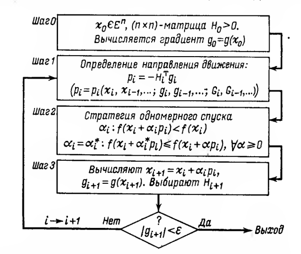


Рис. 4. Структурная схема методов безусловной минимизации.

метод называется методом с памятью. Если же вектор направления p_i явно зависит лишь от величин, вычисленных в точках x_i, то соответствующий метод называется методом без памяти.

В зависимости от максимального порядка производных, входящих в выражение (I,42), алгоритмы минимизации относятся соответственно к методам нулевого, первого и второго порядков. Например, в методах нулевого порядка предусматривается такое построение последовательности $\{p_i\}$, при котором используется лишь информация о значениях минимизируемой функции в различных точках.

Существуют алгоритмы минимизации (нулевого порядка), такие как симплекс- и комплекс-методы, в которых отсутствует построение направлений спуска p_i; они достаточно хорошо освещены в литературе [8, 9] и здесь не приводятся; метод Пауэлла, в котором используется система сопряженных направлений, рассматривается в работе [10], а также [11, с. 121—125].

В методах первого порядка вектор направления обычно определяется из соотношения (I,41). В этом случае последовательность матриц H_i определяется характером применяемого метода, причем в формировании H_i участвуют производные функции $f(x)$ не выше первого порядка. Методы второго порядка допускают зависимость H_i в выражении (I, 42) от вторых частных производных минимизируемой функции. Например, классический метод Ньютона соответствует выбору $H_i = G_i^{-1}$, т. е.

$$p_i = -G_i^{-1} g_i \qquad (I,43)$$

Вектор p_i, $i = 0, 1, \ldots$ обычно определяет направление «спуска», т. е. убывания функции $f(x)$. Это означает, что производная функции $f(x)$ по направлению p_i отрицательна

$$\frac{\partial f(x_i)}{\partial p_i} = p_i^{\mathrm{T}} g_i = -g_i H_i g_i < 0 \qquad (I,44)$$

Достаточным условием выполнения этого условия является положительная определенность матриц H_i, в этом заключается неотъемлемое свойство целого ряда обсуждаемых ниже алгоритмов.

Шаг 2. Рассматривается функция одной переменной

$$h(\alpha) = f(x_i + \alpha p_i) \qquad (I,45)$$

и выбирается величина α_i, определяющая шаг по направлению p_i и удовлетворяющая условию:

$$f(x_i + \alpha_i p_i) < f(x_i) \qquad (I,46)$$

и некоторым другим требованиям. Например, если дополнительное условие для α_i имеет вид

$$f(x_i + \alpha_i p_i) \leqslant f(x_i + \alpha p_i) \quad \forall \alpha \geqslant 0 \qquad (I,47)$$

то говорят о точном нахождении минимума функции $f(x)$ в направлении p_i. Соответствующее α_i будем обозначать α_i^*, а точку $x_i + \alpha_i^* p_i$ назовем оптимальной в направлении p_i. В некоторых случаях полагают $\alpha_i = \text{const}$, если удовлетворяется условие (I,46).

Шаг 3. Рассчитываются точка $x_{i+1} = x_i + \alpha_i p_i$ и (в градиентных методах) вектор $g_{i+1} = g(x_{i+1})$, а затем формируется матрица H_{i+1}. В зависимости от результата проверки условия остановки работа алгоритма либо прекращается (в этом случае x_{i+1} с требуемой точностью является оптимальной точкой), либо осуществляется переход к шагу 1 ($i \rightarrow i + 1$). Условием окончания работы алгоритма обычно является неравенство $|g_{i+1}| < \varepsilon$, где $\varepsilon > 0$ — заданное число, или $|f(x_{i+1}) - f(x_i)| < \varepsilon$. Используются также и другие условия окончания, например $|x_{i+1} - x_i| < \bar{\varepsilon}$ вместе с $|g_{i+1}| < \varepsilon$, и т. д.

Пример алгоритма, обладающего подобной структурой, представляет классический метод наискорейшего спуска (метод Коши), соответствующий выбору в выражении (I,41) $H_i \equiv I_n$, $i = 0, 1, \ldots$, где I_n — единичная ($n \times n$)-матрица. Направлением движения здесь является наискорейшее убывание функции в данной точке:

$$p_i = -g_i \quad i = 0, 1, \ldots \tag{I,48}$$

Точки x_i перехода на следующее направление определяются с помощью условия (I,47), т. е. $\alpha_i = \alpha_i^*$.

Характерная особенность этого метода — простота реализации; однако общеизвестны и его недостатки: метод является линейным — даже при минимизации квадратичной функции процесс поиска ее минимума теоретически бесконечен; для функций с сильно вытянутыми линиями уровня (изолиниями) процесс поиска носит явно выраженный зигзагообразный характер и дает слабое продвижение к минимуму; точное определение минимума практически нереально.

Среди алгоритмов решения задач с ограничениями прежде всего следует отметить методы последовательной безусловной минимизации (см. главу IV). Возросший интерес к этим методам связан, по-видимому, с появлением в последнее десятилетие достаточно эффективных квадратичных методов безусловной минимизации. Суть методов последовательной безусловной минимизации, как известно, заключается в построении на основе минимизируемой функции и функций ограничений некоторого семейства функций, зависящих от параметров. Определяется безусловный минимум (или экстремум) каждой функции этого семейства при фиксированных значениях параметров. Оказывается, что при некоторых условиях последовательность полученных решений задач без ограничений сходится к решению исходной задачи при определенном изменении параметров семейства функций.

Анализ поисковых методов показывает, что в них используются следующие основные алгоритмы (процедуры):

1) расчет целевой функции;
2) расчет производных;
3) линейный поиск по данному направлению;
4) выбор матрицы H;
5) учет ограничений.

Алгоритмы 3, 4 и, как правило, алгоритм 5 характеризуют стратегию поиска.

Вследствие большой размерности, трудоемкости задач оптимизации ХТС требуется высокая эффективность перечисленных основных алгоритмов поисковых методов. Рассмотрим их с этой точки зрения.

Известно, что расчет критерия оптимизации сводится к расчету статического режима системы [3, с. 13]. Повышение эффективности алгоритмов расчета статических режимов процессов достигается применением эффективных методов решения систем нелинейных уравнений, а также использованием методов структурного анализа [1, 3]. Методы решения систем нелинейных уравнений будут подробно рассмотрены в главе II.

Производные критерия оптимизации по варьируемым переменным определяются либо аналитически, либо с помощью разностей:

$$\frac{\partial f}{\partial x_k} \approx \frac{f(x_1, \ldots, x_k + \Delta x_k, \ldots, x_n) - f(x_1, \ldots, x_k, \ldots, x_n)}{\Delta x_k}; \; k=1, \ldots, n \quad (I,49)$$

Важнейшую роль в повышении эффективности поисковых методов играют собственно алгоритмы стратегии поиска (алгоритмы 3, 4, 5). Принципам построения эффективных алгоритмов стратегии поиска уделяется в данной книге основное внимание. Главными критериями оценки эффективности методов минимизации являются необходимая для реализации и работы алгоритма память вычислительной машины, быстродействие алгоритма и надежность определения минимума.

Память, необходимая для реализации алгоритма, зависит от типа используемой ЭВМ, избранного алгоритмического языка программирования, от искусства программирования. Память, необходимая для работы алгоритма определяется, в основном, длиной используемых массивов, зависящих от n — размерности задачи.

Для оценки быстродействия алгоритмов минимизации введем следующие обозначения: τ_f, τ_g, τ_a — время, затраченное соответственно на вычисление значений минимизируемой функции $f(x)$, расчет градиента $g(x)$, работу собственно алгоритма (сумма $\tau_f + \tau_g + \tau_a$ составляет общее время работы алгоритма при решении конкретной задачи). Для реальных задач всегда выполняется условие:

$$\tau_a \ll \tau_f + \tau_g \quad (I,50)$$

Так как τ_f пропорционально общему числу вычислений функции $K(f)$, а $\tau_g \sim K_p$ — числу направлений спуска (предполагается, что алгоритм линейного поиска по направлению требует расчета лишь значений минимизируемой функции в различных точках, а не ее градиента), то быстродействие градиентных алгоритмов минимизации определяется числами K_f и K_p, которые будут в дальнейшем указываться при решении тестовых примеров и конкретных задач. Если производные определяются разностным методом, то общее число вычислений функции равно

$$K = K_f + K_p n \quad (I,51)$$

Качественные оценки надежности и быстродействия алгоритма обычно определяются по результатам минимизации известного на-

бора тестовых функций. Для бо̀льшей части излагаемых алгоритмов в соответствующих местах книги приведены таблицы с этими результатами. Если это специально не оговаривается, производные тестовых функций вычисляются аналитически. Кроме того, для краткости в графах «функция» таблиц приведены порядковые номера тестовых функций, соответствующие приложению 1, а f^* и $|g^*$ обозначают величины f и $|g|$ в конечной точке x^*.

В заключение следует отметить, что в книге рассматриваются только детерминированные локальные методы поиска. Методам случайного поиска посвящена книга [12]. Методы глобального поиска рассматриваются в работе [13, с. 491—525]. Таким образом, в дальнейшем предполагается, что либо множество S выпукло и $f(x)$ выпукла на S, т. е. $f(x)$ имеет в S единственный минимум, либо начальное приближение выбрано достаточно близко от минимума.

Подходы к задаче оптимизации на основе выбора поисковых переменных

Будем исходить из предположения, что ограничения на выходные переменные системы имеют вид (I, 10), а ограничения (I, 11), (I, 12) на переменные состояния отсутствуют. Обозначим через x', z' соответственно векторы промежуточных входных и выходных переменных всех блоков схемы, через $\bar{z}$ — вектор закрепленных [см. (I, 10)] выходных переменных всех блоков схемы, через $\bar{\bar{z}}$ вектор свободных выходных переменных всех блоков, являющихся выходными переменными схемы, через u вектор управлений всех блоков ХТС. Векторы x', z' будут m-векторами, вектор $\bar{z}$ — $\bar{g}$-вектором, вектор $\bar{\bar{z}}$ — $(g - \bar{g})$-вектором, вектор u — $\bar{r}$-вектором, где

$$m = \sum_{k=1}^{N} (m_k - s_k) \qquad \bar{g} = \sum_{k=1}^{N} \bar{g}_k \qquad \bar{g} = \sum_{k=1}^{N} g_k \qquad \bar{r} = \sum_{k=1}^{N} \bar{r}_k \quad (I,52)$$

Тогда совокупность систем уравнений (I, 1) для всех блоков системы может быть записана в виде:

$$\psi'(z', x', u) \equiv z' - f'(x', u) = 0 \qquad (I,53)$$

$$\bar{\psi}(\bar{z}, x', u) \equiv \bar{z} - \bar{f}(x', u) = 0 \qquad (I,54)$$

$$\bar{\bar{\psi}}(\bar{\bar{z}}, x', u) \equiv \bar{\bar{z}} - \bar{\bar{f}}(x', u) = 0 \qquad (I,55)$$

где через f' обозначен вектор правых частей соотношений (I, 1), соответствующих промежуточным переменным всех блоков системы, через $\bar{f}$ вектор правых частей соотношений (I, 1), соответствующих закрепленным выходным переменным всех блоков ХТС, через $\bar{\bar{f}}$ вектор правых частей соотношений (I, 1), соответствующих свободным выходным переменным всех блоков ХТС.

Совокупность соотношений связей для всех блоков системы имеет вид:

$$x' - z' = 0 \qquad (I,56)$$

Предположим, что ограничения на управления заданы в форме неравенств (I, 9). В новых обозначениях совокупность этих ограничений может быть представлена следующим образом:

$$u \in D_1 \qquad D_1 = \{u : c \leqslant u \leqslant b\} \tag{I,57}$$

В новых обозначениях ограничения (I, 10) запишутся так:

$$\bar{z} - a = 0 \tag{I,58}$$

Используя введенные обозначения, запишем критерий оптимизации (I, 15)

$$F = F(u, \bar{\bar{z}}) \tag{I,59}$$

Подставляя в (I, 59) значение $\bar{\bar{z}}$ из (I, 55), приведем критерий к следующему виду:

$$\bar{F}(x', u) \equiv F(u, \bar{\bar{f}}(x', u)) \tag{I,60}$$

В сформулированной выше задаче оптимизации имеются переменные следующих трех типов:

$$x', u, z' \tag{I,61}$$

При постановке любой задачи оптимизации часть переменных (I, 61) (в частном случае все) принимаются в качестве поисковых (независимых), а часть — в качестве зависимых. Поисковыми, или независимыми, называются переменные, в пространстве которых ведется поиск минимального значения критерия (I, 15). Зависимыми переменными являются те из переменных (I, 61), которые на каждом шаге процедуры оптимизации, т. е. при каждом вычислении критерия (I, 15), определяются с помощью систем (I, 53), (I, 54), (I, 56) или их частей для заданных значений независимых переменных. При этом та часть системы (I, 53), (I, 54), (I, 56), которая используется для определения зависимых переменных, будет автоматически удовлетворяться на каждом шаге оптимизации, уравнения же оставшейся части системы (I, 53), (I, 54), (I, 56) необходимо считать ограничениями типа равенств и учитывать с помощью методов условной минимизации. Метод решения задачи оптимизации ХТС существенно зависит от того, какие из переменных (I, 61) будут взяты в качестве поисковых, а какие — в качестве зависимых, какие из уравнений (I, 53), (I, 54), (I, 56), (I, 58) будут удовлетворяться автоматически на каждом шаге оптимизации, а какие необходимо считать ограничениями типа равенств в соответствующей задаче на условный экстремум.

Рассмотрим варианты выбора поисковых переменных.

Поиск в пространстве управлений. В этом случае в качестве поисковых берутся переменные u, а переменные x', z' считаются зависимыми, определяемыми на каждом шаге оптимизации из системы уравнений (I, 53), (I, 56) при заданных значениях поисковых переменных u. Если подставить в уравнение (I, 53) вместо z' вектор x' [в соответствии с (I, 56)], то получим систему уравнений

$$\psi', (x', u) \equiv x' - f'(x', u) = 0 \tag{I,62}$$

С учетом уравнения (I, 54) соотношения (I, 58) примут вид:

$$a - \bar{f}(x', u) = 0 \qquad (I,63)$$

Соотношения (I, 63) принимают во внимание двумя способами.

При первом способе их считают ограничениями типа равенств, которые налагаются на поисковые переменные u и должны учитываться с помощью методов условной минимизации. В этом случае задача оптимизации ХТС выглядит следующим образом (*задача 1*):

$$\min \bar{\bar{F}}(u, x') = 0 \qquad (I,64)$$

$$\psi'(x', u) = 0 \qquad (I,65)$$

$$u \in D_2 \qquad D_2 = \{u : u \in D_1, \quad a - \bar{f}(x', u) = 0\} \qquad (I,66)$$

Итак, поиск ведется в пространстве переменных u; при каждом фиксированном значении u переменные x' находят из системы уравнений (I, 65) [см. выражение (I, 62)].

После того как определены переменные x', переменные z' просто находятся из выражения (I, 56). Таким образом, на каждом шаге оптимизационной процедуры уравнения (I, 65) автоматически удовлетворяются, и с помощью методов условной минимизации необходимо учитывать только ограничения (I, 66).

При втором способе учета ограничений (I, 63) их включают в число уравнений, которые автоматически удовлетворяются на каждом шаге оптимизационной процедуры. Таким образом, число уравнений, которым удовлетворяют переменные u, x', возрастает на величину $\bar{g}$ [см. выражение (I, 52)] по сравнению с предыдущим случаем. На такую же величину должно возрасти число зависимых переменных, с помощью которых должна удовлетворяться система уравнений (I, 62), (I,63). Предположим, что последние $\bar{s}_k$ ($\bar{s}_k \leqslant s_k$) управляющих переменных в каждом блоке, т. е. часть свободных входных переменных блока [см. выражение (I, 13)] будут включены в число зависимых переменных. Введем обозначение

$$\bar{s} = \sum_{k=1}^{N} \bar{s}_k \qquad (I,67)$$

Тогда должно существовать равенство

$$\bar{s} = \bar{g} \qquad (I,68)$$

Обозначим через $\bar{u}$ вектор, компонентами которого будут

$$u_i^{(k)} \qquad i = 1, \ldots, r_k, r_k + 1, \ldots, r_k + s_k - \bar{s}_k \qquad k = \overline{1, N} \qquad (I,69)$$

а через $\bar{\bar{u}}$ — вектор, компонентами которого будут

$$u_i^{(k)} \qquad i = r_k + s_k - \bar{s}_k + 1, \ldots, r_k + s_k \qquad (I,70)$$

Тогда независимыми окажутся переменные $\bar{u}$, а зависимыми — переменные x', $\bar{\bar{u}}$ и задача оптимизации ХТС формулируется следующим образом (*задача 2*):

$$\min_{\bar{u}} \bar{\bar{F}}(u,\ x') \tag{I,71}$$

$$\psi'(x',\ \bar{u},\ \bar{\bar{u}}) = 0 \tag{I,72}$$

$$a - \bar{f}(x',\ \bar{u},\ \bar{\bar{u}}) = 0 \tag{I,73}$$

$$u \in D_1 \tag{I,74}$$

В данном случае запись вектор-функций ψ', $\bar{f}$ отличается от записи при формулировке задачи 1 [см. выражения (I, 65), (I, 66)] тем, что здесь отражено разбиение вектора u на два вектора: $\bar{u}$, $\bar{\bar{u}}$. Итак, поиск ведется в пространстве независимых переменных $\bar{u}$; при каждом фиксированном значении $\bar{u}$ зависимые переменные x', $\bar{\bar{u}}$ находятся из системы уравнений (I, 72), (I, 73). Таким образом, на каждом шаге оптимизационной процедуры автоматически удовлетворяются уравнения (I, 72), (I, 73), и с помощью методов условной минимизации необходимо учитывать только простые ограничения (I, 74).

Решение системы уравнений (I, 1), (I, 2) [или эквивалентной ей системы (I, 53), (I, 54)] при фиксированных значениях управляющих переменных u является по существу расчетом стационарного режима ХТС (расчетом материального и теплового баланса системы). Для простоты будем называть решение системы (I, 65) расчетом системы. Решение же системы (I, 72), (I, 73) является по существу расчетом системы при одновременном удовлетворении некоторых граничных условий.

Решение системы (I, 65) иногда называют моделирующим расчетом схемы, а решение системы (I, 72), (I, 73) — проектным расчетом схемы.

Поиск в пространстве управлений и переменных состояния. В этом случае независимыми являются переменные x', u, а зависимыми z'. Для определения z' будут использоваться уравнения (I, 56). Исключив переменные z' из выражения (I, 53) с помощью (I, 56), получим систему уравнений (I, 62). Задача оптимизации ХТС записывается следующим образом (*задача 3*):

$$\min \bar{\bar{F}}(u,\ x') \tag{I,75}$$

$$x',\ u \in D_3 \tag{I,76}$$

$$D_3 = \{x',\ u : x' - f'(x',\ u) = 0,\quad \bar{f}(x',\ u) - a = 0,\quad u \in D_1\}$$

Следовательно, поиск ведется в пространстве переменных x', u а с помощью методов условной минимизации необходимо добиваться удовлетворения как системы (I, 62), так и системы (I, 63), которые во время поиска рассматриваются как ограничения типа равенства.

В ряде случаев в качестве независимых переменных удобно выбрать управления u и часть переменных состояния x'; тогда зависимыми переменными будут оставшаяся часть переменных x', а также переменные z'. Обозначим через $\bar{x}'$ вектор зависимых переменных состояния, а через $\bar{\bar{x}}''$ — вектор независимых переменных

состояния. Так же как и во всех предыдущих случаях, исключив переменные z' из уравнений (I, 53), с помощью уравнения (I, 56) получим систему (I, 62), которую запишем в виде:

$$\bar{x}' - \bar{f}'(\bar{x}', \bar{\bar{x}}' u) = 0 \qquad (I,77)$$

$$\bar{\bar{x}}' - \bar{\bar{f}}'(\bar{x}', \bar{\bar{x}}', u) = 0 \qquad (I,78)$$

где через $\bar{f}'$ обозначена часть вектора функции f', соответствующая переменным $\bar{x}'$, а через $\bar{\bar{f}}'$ — часть вектора f', соответствующая переменным $\bar{\bar{x}}'$. Система уравнений (I, 77) будет автоматически удовлетворяться на каждом шаге оптимизационной процедуры, а уравнения системы (I, 78) будут считаться ограничениями типа равенств, которые должны учитываться с помощью методов условной минимизации. В этом случае задача оптимизации ХТС выглядит следующим образом (*задача 4*):

$$\min_{u, \bar{\bar{x}}'} F(u, \bar{x}', \bar{\bar{x}}') \qquad (I,79)$$

$$\bar{x}' - \bar{f}'(\bar{x}', \bar{\bar{x}}', u) = 0 \qquad (I,80)$$

$$u, \bar{\bar{x}}' \in D_4 \qquad (I,81)$$

$$D_4 = \{u, \bar{\bar{x}}' : u \in D_1,\ a - \bar{f}(x', u) = 0,\ \bar{\bar{x}}' - f'(x', u) = 0\}$$

Итак, в данном случае поиск ведется в пространстве переменных u, $\bar{\bar{x}}'$; при каждом фиксированном значении переменных u, $\bar{\bar{x}}'$ переменные $\bar{x}'$ находятся из системы уравнений (I, 80), а ограничение (I, 81) удовлетворяются с помощью методов условной минимизации. Преимущества и недостатки сведения задачи оптимизации ХТС к сформулированным задачам будут рассмотрены нами в гл. IV после изучения методов условной и безусловной минимизации. Здесь мы только отметим, что в большинстве случаев задача оптимизации ХТС сводится либо к задаче 1, либо (реже) к задаче 4.

Дадим схемную интерпретацию перехода от задачи 1 к задаче 3 или 4. Пусть в начале задача оптимизации ХТС сформулирована как задача 1. Выберем в качестве независимых переменных помимо управлений $u^{(k)}$, $(k = \overline{1, N})$ также входные промежуточные перемены $x_i^{(p)}$, $(i = s_p + 1, \ldots, n_p)$ p-го блока. В этом случае из системы уравнений (I, 53), (I, 56) или эквивалентной ей системы (I, 65) [либо (I, 62)], которая автоматически удовлетворялась при каждом оптимизационном шаге (шаге второго уровня), исключается совокупность уравнений

$$x_i^{(p)} - z_{g_{p,i}}^{h_{p,i}} = 0 \qquad i = s_p + 1, \ldots, n_p \qquad (I,82)$$

и переносится в совокупность уравнений, которые удовлетворяются с помощью методов условной минимизации. Поскольку соотношения (I, 82) характеризуют все потоки, которые подаются на вход k-го

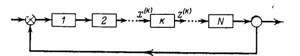
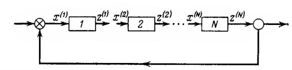


Рис. 5. Схема последовательности блоков с рециклом.

Рнс. 6. Разомкнутая схема, соответствующая схеме, приведенной на рис. 5.



блока, исключение соотношений (I, 82) из системы (I, 53) эквивалентно разрыву всех входных потоков p-го блока.

Итак, со схемной точки зрения выбор входных промежуточных переменных p-го блока в качестве независимых переменных эквивалентен разрыву всех входных потоков p-го блока. При этом выходные переменные блоков, потоки из которых подаются на вход p-го блока, становятся свободными выходными переменными схемы (зависимыми переменными). Следовательно, переход от задачи 1 к задаче 3 может быть представлен таким образом: разрываются все входные потоки всех блоков схемы, при этом входные промежуточные переменные блоков схемы становятся поисковыми (независимыми) на уровне задачи оптимизации, а выходные промежуточные переменные блоков становятся свободными выходными переменными схемы (зависимыми переменными).

При переходе от задачи 1 к задаче 4 разрывается только часть потоков. Иногда этот прием называют вынесением соотношений связи в критерий [11, с. 181]. На рис. 5 приведена схема с рециклом, а на рис. 6 — соответствующая разомкнутая схема, полученная при выборе входных переменных 2-го блока в качестве независимых.

ГЛАВА II

Расчет стационарных режимов химико-технологических систем

Пусть химико-технологическая система состоит из N блоков. Для простоты записи будем исходить из предположения, что каждый блок имеет один входной и один выходной поток размерности m. В отличие от формы математической модели, принятой нами при постановке задачи оптимизации ХТС [см. выражение (I, 1)], запишем математическую модель k-го блока в более общем виде

$$\bar{\varphi}^{(k)}(v^{(k)}, w^{(k)}, u^{(k)}) = 0 \qquad (II,1)$$

где

$$v^{(k)} = \left(v_1^{(k)}, \ldots, v_m^{(k)}\right) \qquad w^{(k)} = \left(w_1^{(k)}, \ldots, w_m^{(k)}\right)$$

m — векторы, характеризующие соответственно входной и выходной физический поток; $u^{(k)}$ — r_k-вектор управляющих переменных блоков (в данной главе будем считать его заданным): $u^{(k)} = \left(u_1^{(k)}, \ldots, u_{r_k}^{(k)}\right)$; $\bar{\varphi}^{(k)}\left(\bar{\varphi}_1^{(k)}, \ldots, \bar{\varphi}_m^{(k)}\right)$ — вектор левых частей системы (II, 1).

Модель (II, 1) относительно выходных переменных записана в неявном виде, поскольку для ряда аппаратов (реактор идеального смешения, абсорбер и др.) выходные переменные действительно являются неявными функциями входных переменных. Выражение (II, 1) представляет собой систему из m уравнений с $2m$ неизвестными. Если задать любые m чисел $v^{(k)}$ или $w^{(k)}$, или часть переменных $v^{(k)}$ и часть $w^{(k)}$, то, вообще говоря, система (II, 1) позволяет найти остальные m чисел. В дальнейшем, в отличие от физических входных $v^{(k)}$ и выходных $w^{(k)}$ переменных блока введем расчетные переменные: входные $x^{(k)}$ (при расчете блока считаются известными) и выходные $z^{(k)}$ (получаются в результате расчета блока). Это связано с тем, что при расчете схемы направление расчета блока не всегда совпадает с направлением физических потоков, входящих и выходящих из блока. Иногда выбор того или иного направления расчета блока может существенно упростить его расчет [3, с. 24].

Соотношения связи блоков записывают следующим образом:

$$v^{(k)} = \sum_{j=1}^{N} \alpha_{kj} w^{(j)} \qquad \left(\sum_{k=1}^{N} \alpha_{kj} = 1 \right) \qquad k = \overline{1, N} \tag{II,2}$$

Если расчетные входные и выходные переменные блоков выбраны, то математические модели блоков описывают выражениями (I, 1), а соотношения связи — равенствами (I, 6). В частном случае модель k-го блока может быть линейна:

$$z^{(k)} = A^{(k)} x^{(k)} + в^{(k)} \tag{II,3}$$

В дальнейшем блок-схему химико-технологического процесса будем называть технологической, если направления ее потоков совпадают с направлениями физических потоков, и информационной, если направления ее потоков совпадают с направлениями расчета блоков.

Рассмотрим два типа расчетов — моделирующий и так называемый проектный.

Моделирующий расчет. В этом случае все входные переменные схемы и управляющие переменные блоков считаются заданными, а выходные переменные схемы — свободными, так что задача расчета стационарного режима системы сводится к решению системы из $2Nm$ нелинейных уравнений (II, 1), (II, 2) с $2Nm$ неизвестными $v^{(k)}$, $w^{(k)}$ ($k = \overline{1, N}$). Различают два подхода к расчету схем — модульный [14, с. 15—18] и ориентированный на уравнения [14, с. 23]. При модульном подходе к системе уравнений (II, 1) любого блока ХТС подходят как к неразделимому целому, другими словами, не разрешается обрабатывать отдельные уравнения системы (II, 1). Это значит, что блок начинают рассчитывать только, когда задан любой набор из m переменных $v^{(k)}$, $w^{(k)}$, после чего одновременно вычисляют оставшиеся m переменных. При применении ориентированного на уравнения подхода уравнения всех блоков ХТС собираются в единую систему уравнений, каждое из которых может обрабатываться отдельно. Это значит, что по заданным всем (или

части) компонентам вектора $v^{(k)}$ разрешается рассчитывать одно или несколько $w_i^{(k)}$, или наоборот. При модульном подходе возможны совместный (параллельный) и последовательный методы расчета ХТС. В свою очередь параллельные методы делятся на одноуровневые и двухуровневые [15, 16]. При одноуровневом параллельном методе расчета все переменные $v^{(k)}$ и $w^{(k)}$ считаются итерируемыми, и мы должны решать систему из $2Nm$ нелинейных уравнений (II, 1), (II, 2) с $2Nm$ неизвестными $v^{(k)}$, $w^{(k)}$ $(k = \overline{1, N})$. В том случае, когда в каждом блоке определены расчетные входные и выходные переменные, выходные переменные блоков могут быть исключены. Действительно, подставив в равенство (I, 6) уравнения (I, 1), получим систему из Nm уравнений

$$F(x) \equiv x^{(k)} - \sum_{j=1}^{N} \alpha_{kj} f^j (x^{(j)}, u^{(j)}) = 0 \qquad k = \overline{1, N} \qquad \text{(II,4)}$$

относительно $\overline{N} = mN$ неизвестных $x^{(k)}$ $(k = \overline{1, N})$.

При двухуровневых параллельных методах все переменные $v^{(k)}$, $w^{(k)}$ также считаются итерируемыми, при этом на k-том шаге итерации проводится линеаризация моделей (II, 1), для чего используется их специальный вид. После этого система уравнений (II, 1), (II, 3) становится линейной и ее решают одним из известных методов. В результате решения мы получаем новую точку, в которой опять проводится линеаризация моделей и т. д. Часто эти методы оказываются весьма эффективными. Однако они не универсальны, поскольку обычно в них используют специальный вид моделей блоков (II, 1).

Рассмотрим последовательный метод расчета. Он применяется к информационной блок-схеме. В связи с этим необходим этап преобразования технологической блок-схемы в информационную. В простейшем случае информационная блок-схема строится таким образом, что направления информационных потоков совпадают с направлениями соответствующих физических потоков. Однако при рациональном выборе направлений потоков иногда удается либо уменьшить размерность решаемой системы нелинейных уравнений, либо вообще перейти к безытерационному расчету.

После построения информационной блок-схемы переходят к выбору итерируемых переменных; в качестве таковых берут не все переменные $x^{(k)}$, $z^{(k)}$, как при параллельном методе, а только часть. Для определенности будем исходить из предположения, что в качестве итерируемых выбраны компоненты некоторой совокупности векторов

$$x^{(k)} \qquad (k = p_1, \ldots, p_s) \qquad s < N \qquad \text{(II,5)}$$

Совокупность векторов (II, 5) обладает тем свойством, что разрыв потоков, соответствующих этим векторам, превращает замкнутую схему в разомкнутую.

В дальнейшем разрываемые потоки будем называть особыми, а схему, полученную разрывом особых потоков — соответствующей разомкнутой схемой.

После того, как переменные (II, 5) выбраны, расчет ХТС сводится к решению некоторой системы нелинейных уравнений вида:

$$x = F(x, u) \qquad (II,6)$$

где x — вектор размерности $n = ms$, компонентами которого являются все компоненты векторов (II, 5); вектор $F(x, u) = [F_1(x, u), \ldots, F_n(x, u)]$. При заданных значениях x, u вектор $F(x, u)$ определяется расчетом соответствующей разомкнутой схемы [1, с. 30]. Аналитический вид функции $F(x, u)$ нам неизвестен. Иногда систему (II, 6) удобно представить так:

$$f(x, u) = 0 \qquad [f(x, u) \equiv x - F(x, u)] \qquad (II,7)$$

Проиллюстрируем различные подходы к расчету ХТС на примере простой последовательности блоков с рециклом (см. рис. 5). В случае параллельного расчета в качестве итерируемых переменных могут быть взяты либо все входные $x^{(k)}$ $(k = \overline{1, N})$ и выходные $z^{(k)}$ $(k = \overline{1, N})$ переменные блоков, либо только входные переменные. Рассмотрим теперь последовательный метод расчета. Пусть направления потоков в информационной схеме совпадают с направлениями физических потоков. В этом случае вид информационной блок-схемы также соответствует схеме, приведенной на рис. 5. В качестве итерируемых могут быть взяты, например, переменные $x^{(2)}$ (либо переменные, соответствующие любому потоку, связывающему блоки 1, ..., N). Соответствующая разомкнутая схема приведена на рис. 6.

В настоящее время в большинстве случаев применяется модульный последовательный метод расчета ХТС. В связи с этим все дальнейшее изложение будем вести применительно к этому методу (если не будет оговариваться противное), называя его для простоты последовательным методом расчета ХТС.

Проектный расчет. Остановимся коротко на проектном расчете ХТС. В этом случае уравнения математической модели ХТС дополняются уравнениями, выражающими ограничения на выходные переменные ХТС, и полученная система уравнений решается совместно. Пусть ограничения на выходные переменные имеют вид (I, 10). Тогда уравнения (I, 10) добавляются к системе (I, 1), (I, 2) или к эквивалентной ей системе (II, 7). Поскольку число уравнений увеличилось на величину $\bar{g}$ [см. выражение (I, 52)], на столько же должно увеличиться число неизвестных, куда дополнительно должны быть включены $\bar{g}$ управляющих или входных переменных схемы (при моделирующем расчете они считались заданными). Для определенности будем считать, что в число неизвестных включаются управляющие переменные (I, 13) (часть свободных входных переменных ХТС), причем соблюдается равенство (I, 68). П е р в ы й п у т ь решения полученной системы нелинейных уравнений состоит в организации двухуровневого расчета, когда на первом уровне решаются уравнения (II, 7) при фиксированных значениях переменных (I, 13), а на втором идет такой подбор переменных (I, 13), при котором удовлетворялись бы условия (I, 10). Подобный принцип проектного расчета принят во многих системах программ моделиро-

вания ХТС, например в работе [19]. Ясно, что при этом на первом уровне проводится моделирующий расчет ХТС. Второй путь состоит в организации одноуровневого расчета, когда одновременно решается система уравнений (II, 7), (I, 10) относительно неизвестных x и u [см. формулы (I, 13)]. В этом случае в информационную блок-схему искусственно вводятся дополнительные обратные связи; подробнее об этом см. в работах [1, с. 28; 30].

Заметим, что моделирующий расчет ХТС соответствует решению системы (I, 62) при фиксированных значениях управлений u, а проектный — решению систем (I, 72), (I, 73) относительно переменных x', $\bar{\bar{u}}$ при фиксированных значениях переменных $\bar{u}$.

Итак, расчет стационарного режима ХТС сводится к решению некоторой системы нелинейных уравнений. Поэтому все дальнейшее изложение будет посвящено методам решения систем нелинейных уравнений. Заметим, что имеется определенная специфика решения систем нелинейных уравнений при использовании последовательного подхода. Действительно, при заданном x мы не можем рассчитать отдельно левую часть одного или нескольких уравнений системы (II,7), рассчитать их можно только вместе. Это не позволяет использовать методы, в которых предусмотрена обработка каждого уравнения системы (II, 7) в отдельности (например, метод Гаусса—Зейделя [20, с. 345] в случае линейных систем, метод Брауна [21] в случае нелинейных систем).

Методы решения систем нелинейных уравнений можно разбить на три группы. К первой относятся метод простой итерации и его модификации, а также методы, ускоряющие сходимость простой итерации (методы *DEM* [22], *GDEM* [23]); ко второй — метод Вольфа и его модификации [3, с. 35; 1, с. 84]; к третьей — квазиньютоновские методы. Здесь мы рассмотрим только метод Ньютона и квазиньютоновские методы решения систем нелинейных уравнений, идейно очень близкие к методу Ньютона и квазиньютоновским методам оптимизации. В дальнейшем будем говорить, что метод обладает *p-шаговым свойством линейного окончания*, если он обеспечивает решение системы линейных уравнений при числе шагов, не превышающем p.

Отметим, что к задаче решения систем нелинейных уравнений помимо расчета стационарных режимов ХТС сводится и ряд других важных задач в химической технологии — например, расчет противоточных аппаратов (ректификационные колонны, абсорберы, теплообменники и др.), поэтому она является одной из наиболее распространенных вычислительных задач в химической технологии.

Метод Ньютона

Пусть необходимо решить систему нелинейных уравнений

$$f(x) = 0 \qquad (II,8)$$

где f, x — n-векторы. Обозначим через x_j ($j = 1, \ldots, n, \ldots$) последовательность точек, полученных при использовании метода Нью-

тона. Элементы вектора x_j будем обозначать через x_{ij} ($i = \overline{1, n}$). Разложим в точке x_j левую часть системы уравнений (II, 8) в ряд Тейлора и оставим члены порядка малости не выше первого

$$f(x_j + \Delta x_j) = f(x_j) + J_j \Delta x_j \qquad (II,9)$$

где J_j — матрица Якоби системы функции f; Δx_j — приращения, которые получают переменные x_j на i-том шаге метода Ньютона. Введем обозначения:

$$f_j = f(x_j) \qquad y_j = f_{j+1} - f_j \qquad s_j = x_{j+1} - x_j \qquad (II,10)$$

и найдем величины Δx_j. Для этого приравняем правую часть равенства (II, 9) нулю, считая точку x_{j+1} корнем уравнения (II, 8) в линейном приближении:

$$J_j \Delta x_j = - f_j \qquad (II,11)$$

отсюда имеем

$$\Delta x_j = - J_j^{-1} f_j \qquad (II,12)$$

Чтобы найти следующее приближение

$$x_{j+1} = x_j + \Delta x_j \qquad (II,13)$$

необходимо определить вектор Δx_j либо как решение системы линейных уравнений (II, 11), либо с помощью матрицы, обратной матрице Якоби [см. выражение (II, 12)]. Для повышения стабильности метода следующее приближение часто находят с помощью соотношения

$$x_{j+1} = x_j + \alpha_j p_j \qquad (II,14)$$

где α_j — скаляр, а вектор p_j определяется либо решением системы уравнений

$$J_j p_j = - f_j \qquad (II,15)$$

либо с помощью соотношения

$$p_j = - J_j^{-1} f_j \qquad (II,16)$$

Вектор p_j равен ньютоновскому шагу в j-той точке. Ясно, что $s_j = \alpha_j p_j$. Введем обозначение

$$N_j = N(x_j) = \sqrt{\sum_{i=1}^{n} f_{ij}^2} \qquad (II,17)$$

Тогда параметр α_j выбирается либо из условия

$$\alpha_j = \arg\min_{\alpha} N(x_j + \alpha p_j) \qquad (II,18)$$

либо из условия

$$N_{j+1} < N_j \qquad (II,19)$$

Легко показать, что метод Ньютона позволяет находить решение системы линейных уравнений за один шаг. Действительно, пусть выражение (II, 8) является системой линейных уравнений, т. е.

$$f(x) \equiv Ax + b = 0 \qquad (II,20)$$

где A — невырожденная $n \times n$-матрица; b — n-вектор.

Решение этой системы имеет вид $x^* = -A^{-1}b$. Поскольку

$$\Delta x_j = -A^{-1}(Ax_j + b) = -x_j - A^{-1}b$$

то

$$x_1 = x_0 - x_0 - A^{-1}b = -A^{-1}b$$

Преимущества и недостатки метода Ньютона применительно к задаче оптимизации рассмотрены в работе [11, с. 268]; остановимся на наиболее существенном недостатке. Метод Ньютона требует определения матрицы Якоби — левых частей системы уравнений (II, 8). В случае расчета стационарных режимов ХТС аналитическое определение матрицы Якоби обычно требует очень трудоемкой подготовительной работы. Конечно, положение изменится, когда будут созданы системы программ моделирования ХТС, использующие математический аппарат «сопряженного процесса» [1, с. 139], позволяющий вычислять требуемые производные. Однако, поскольку таких программ, полностью автоматизирующих аналитическое определение матрицы Якоби, пока еще нет, метод Ньютона с аналитическим вычислением производных применяется очень редко. В связи с этим ставится задача использования метода Ньютона с некоторой аппроксимацией матрицы Якоби. Наиболее простым способом получения аппроксимации матрицы Якоби является разностный. В этом случае элементы j_{kp} матрицы J_l подсчитываются следующим образом:

$$j_{kp} = \frac{f_{kl}(x_{1l}, \ldots, x_{pl} + \Delta x_{pl}, \ldots, x_{nl}) - f_{kl}(x_{1l}, \ldots, x_{pl}, \ldots, x_{nl})}{\Delta x_{pl}} \qquad \text{(II,21)}$$

В методе Ньютона с разностной аппроксимацией матрицы Якоби можно выделить два этапа на каждом шаге — это сбор информации для построения аппроксимации матрицы Якоби и собственно поиск. На первом этапе поочередно даются приращения всем аргументам функции $f(x)$, причем функция $f(x)$ вычисляется в $(n + 1)$ точках, и вычисляются элементы матрицы Якоби с помощью уравнения (II, 21). Второй этап — это собственно поиск, при котором в начале определяется направление p_j поиска с помощью выражения (II, 15), а затем делается шаг в этом направлении. Благодаря тому, что эти два этапа разделены, на каждом шаге приходится вычислять функции $f(x)$ в $(n + 1)$-й точке. В этом заключен большой недостаток метода, особенно существенный, когда размерность системы (II, 8) велика.

Квазиньютоновские методы

Основной идеей квазиньютоновских методов является объединение этапов сбора информации и поиска. Причем информация, которую получают во время поиска, используется для построения аппроксимации B_j матрицы Якоби J_j либо аппроксимации H_j матрицы, обратной к матрице Якоби. По аналогии с соотношениями (II, 15), (II, 16) направление поиска определяется либо решением системы линейных уравнений

$$B_j p_j = -f_j \qquad \text{(II,22)}$$

либо с помощью соотношения

$$p_j = -H_j f_j \qquad \text{(II,23)}$$

Следующая точка итерации определяется с помощью формулы (II, 14). Преимущество аппроксимации обратной матрицы Якоби состоит в том, что в этом случае не нужно решать систему линейных уравнений. Однако аппроксимация самой матрицы Якоби имеет свои преимущества, которые мы обсудим ниже. Конечно, информация относительно функции $f(x)$, получаемая во время поиска и используемая для построения матриц B_j, H_j, должна быть достаточно «качественной». Ясно, что если точки поиска x_j достаточно долго будут находиться либо в гиперплоскости, либо в близкой к ней окрестности, то построить аппроксимацию матрицы Якоби будет трудно. Можно отметить некоторую аналогию с методами активного и пассивного эксперимента в теории планирования эксперимента. В методах активного эксперимента для построения математической модели объекта используются специальные возмущения, наносимые на объект. Для построения же математической модели с помощью методов пассивного эксперимента оперируют данными нормальной эксплуатации объекта.

Найдем соотношения, которые должны удовлетворять матрицы B_j и H_j. Перепишем соотношение (II, 9), используя обозначения (II, 10):

$$J_j s_j = y_j \qquad (II,24)$$

Итак, в линейном приближении матрица Якоби удовлетворяет равенству (II, 24). Потребуем, чтобы матрицы B_{j+1} удовлетворяли соотношению, которое аналогично тому, которому удовлетворяет сама матрица Якоби [24]

$$B_{j+1} s_j = y_j \qquad (II,25)$$

Выпишем соотношение (II, 24) для всех предыдущих точек, начиная с нулевой $x_0, \ldots, x_{j-1}$ ($j \leqslant n$). При этом предположим, что точки $x_0, \ldots, x_{j-1}$ находятся вблизи решения системы (II, 8) и что в его окрестности эта система близка к линейной. При этом предположении элементы матрицы Якоби можно считать постоянными, не зависящими от номера точки. Отсюда равенства (II, 24) для $j = 0, 1, \ldots, i-1$ ($i \leqslant n$) примут вид

$$J s_j = y_j \qquad j = 0, 1, \ldots, i-1 \qquad i \leqslant n \qquad (II,26)$$

Матрицы Y_i, S_i размерности $n \times i$ введем следующим образом:

$$Y_i = (y_0, \ldots, y_{i-1}) \qquad S_i = (s_0, \ldots, s_{i-1}) \qquad (II,27)$$

Тогда совокупность систем уравнений (II, 26) будет эквивалентна следующему матричному уравнению:

$$JS_i = Y_i \qquad i \leqslant n \qquad (II,28)$$

Можно потребовать, чтобы на i-том шаге искомая матрица B_i удовлетворяла уравнению

$$B_i S_i = Y_i \qquad i \leqslant n \qquad (II,29)$$

которое аналогично уравнению (II, 28). В системе линейных уравнений (II, 29) при $i < n$ число неизвестных (элементы матрицы B_i) больше числа уравнений, при $i = n$ оно становится равным числу уравнений. Далее будет показано, что в случае линейных систем

определение матрицы B_i из уравнения (II, 29) обеспечивает решение этой системы на $(n+1)$-м шаге. Это, собственно, и оправдывает введение этого уравнения. В случае нелинейных систем матричное уравнение

$$B_i S_i = Y_i \qquad i > n$$

уже не имеет решения, поскольку число неизвестных в нем меньше числа уравнений. Способы продолжения процедуры решения при $i > n$ будут обсуждены ниже.

Рассмотрим теперь случай, когда строится приближение к обратной матрице Якоби. Из (II, 24) имеем

$$J_j^{-1} y_j = s_j \tag{II,30}$$

Потребуем, чтобы приближение H_{j+1} к матрице J_{j+1}^{-1} удовлетворяло условию

$$H_{j+1} y_j = s_j \tag{II,31}$$

Если потребовать, чтобы матрица H_i удовлетворяла соотношениям (II, 31) в i предыдущих точках, начиная с нулевой, то по аналогии с уравнением (II, 29) можно получить следующее матричное уравнение для определения H_i [25; 26]:

$$H_i Y_i = S_i \qquad i \leqslant n \tag{II,32}$$

Процедура определения матрицы H_i при $i > n$ будет обсуждена ниже.

В дальнейшем соотношения (II, 25), (II, 31) будем называть квазиньютоновскими условиями 1-го рода. Соответствующие методы, в которых матрицы B_i, H_i удовлетворяют этим условиям, будут называться квазиньютоновскими методами 1-го рода. Соотношения (II, 29), (II, 32) будем называть квазиньютоновскими условиями 2-го рода, а соответствующие методы — квазиньютоновскими методами 2-го рода. Будем говорить, что квазиньютоновский метод обладает глубиной памяти q, если в $(k+1)$-й точке матрицы B_{k+1}, H_{k+1} должны удовлетворять квазиньютоновским условиям в q предыдущих точках:

$$B_{k+1} s_j = y_j \qquad q \leqslant k \; j = k, \qquad k-1, \ldots, k-q+1 \tag{II,33}$$

$$H_{k+1} y_j = s_j \qquad q \leqslant k \; j = k, \qquad k-1, \ldots, k-q+1 \tag{II,34}$$

Таким образом, квазиньютоновские методы 1-го рода являются методами с глубиной памяти, равной 1, и для построения матрицы Якоби (обратной матрицы Якоби) в $(j+1)$-й точке они используют только информацию в данной точке [векторы s_j, y_j в соотношениях (II, 25), (II, 31)]. В то же время для построения матриц B_j, H_j методы с глубиной памяти q используют и предыдущую информацию. Методы 2-го рода отличаются тем, что глубина памяти q увеличивается в них на 1 на каждом шаге (при $i \leqslant n$).

Рассмотрим частный случай, когда система (II, 8) является системой линейных уравнений (II, 20). Условия (II, 24) в данном случае запишутся в виде:

$$A s_j = y_j \qquad j = \overline{0, n-1} \tag{II,35}$$

а уравнения (II,29) — в виде:

$$AS_i = Y_i \tag{II,36}$$

или

$$S_i = A^{-1}Y_i \tag{II,37}$$

Для вывода квазиньютоновских методов здесь будут использованы вариационные методы и аппарат псевдообратных матриц.

Квазиньютоновские методы 1-го рода

Вначале найдем приближение B_{j+1} к самой матрице Якоби. Удобнее искать матрицу B_{j+1} в виде

$$B_{j+1} = B_j + E \tag{II,38}$$

где B_j известна из предыдущей итерации, а матрицу E требуется найти. Подставив B_{j+1} из формулы (II, 38) в соотношение (II, 25), найдем, что матрица E, содержащая n^2 элементов e_{ij}, должна удовлетворять следующему матричному уравнению:

$$Es_j = r_j \tag{II,39}$$

где

$$r_j = y_j - B_j s_j \tag{II,40}$$

Полученное выражение эквивалентно n линейным уравнениям. Таким образом, число неизвестных превышает число уравнений. Имеющимися степенями свободы мы можем воспользоваться, выбрав элементы матрицы E наилучшими с точки зрения какого-либо критерия. Очень важен выбор критерия, который, с одной стороны, должен приводить к хорошей сходимости метода, а с другой, обеспечивать не слишком большую сложность получающихся методов. В последнее время в качестве такого критерия используется минимум нормы Фробениуса матрицы E [27; 28]:

$$\| E \|_F = \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}^2 \tag{II,41}$$

Возникает вопрос — почему взят такой критерий? Можно ли строго доказать, что надо использовать именно этот критерий, а не другой? На это может быть дан только один ответ — отрицательный: не может быть строгого обоснования выбора в качестве критерия нормы Фробениуса. Однако в качестве обоснования обычно приводят следующие качественные соображения. Минимизация критерия (II, 41) обеспечивает наименьшее из возможных изменение матрицы B_j при наложении некоторых дополнительных ограничений, т. е. использование такого критерия обеспечивает максимальную близость матрицы B_{j+1} к матрице B_j. Отсюда, если B_j обладала какими-либо хорошими свойствами (например, была близка к матрице Якоби), то матрица B_{j+1} должна в какой-то степени их сохранить. Конечно, приведенные рассуждения ни в коем случае не являются строгим обоснованием выбора критерия (II, 41). Единственным действительным обоснованием может служить эффективность тех алгоритмов, которые могут быть получены на основе этого критерия (II,41), т. е. только вычислительная практика. Не исключено, что практика подскажет

другой критерий, который даст возможность получать более эффективные методы. Здесь имеется некоторая аналогия с естественными науками, где на основании практики часто выдвигается та или иная гипотеза, следствия которой проверяются практикой же.

В дальнейшем назовем *принципом наименьшего изменения матрицы* B_j способ определения матрицы B_{j+1} из условия минимума нормы Фробениуса при наличии каких-либо дополнительных условий, зависящих от конкретной задачи.

Займемся теперь определением вида матрицы E. Согласно сказанному выше, ее элементы e_{ij} должны находиться из условий выполнения соотношения (II, 39) и минимума критерия (II, 41). Математически эта экстремальная задача запишется следующим образом:

$$\min_{e_{ij}} \frac{1}{2} \| E \|_F \tag{II,42}$$

$$Es_j = r_j \tag{II,43}$$

Перепишем эту задачу в координатном виде:

$$\min_{e_{ij}} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}^2 \qquad \sum_{l=1}^{n} e_{kl} s_{lj} = r_{k,j} \tag{II,44}$$

Для ее решения воспользуемся методом множителей Лагранжа. Функция Лагранжа будет иметь вид:

$$L = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} e_{kl}^2 + \sum_{k=1}^{n} \lambda_k \sum_{l=1}^{n} e_{kl} s_{l,j}$$

где λ_k — множители Лагранжа. Приравнивая производные L по e_{ip} нулю, получим

$$e_{ip} = -\lambda_i s_{p,j}$$

Отсюда легко проверить, что имеет место матричное соотношение

$$E = -\lambda s_j^{\mathrm{T}} \tag{II,45}$$

где $\boldsymbol{\lambda}$ — вектор с компонентами $\lambda_1, \ldots, \lambda_n$.

Для определения $\boldsymbol{\lambda}$ подставим это выражение для E в уравнение (II, 43), откуда найдем $\boldsymbol{\lambda} = -r_j/(s_j^{\mathrm{T}} s_j)$; следовательно, матрица E будет иметь вид

$$E = \frac{r_j s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} s_j} \tag{II,46}$$

Подставляя выражение для E в (II, 38) и используя (II, 40), получим

$$B_{j+1} = B_j + \frac{(y_j - B_j s_j) s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} s_j} \tag{II,47}$$

Получим теперь выражение для аппроксимации обратной матрицы H_{j+1}. Известна формула Хаусхолдера [24]

$$(A + vw^{\mathrm{T}})^{-1} = A^{-1} - \frac{A^{-1} v w^{\mathrm{T}} A^{-1}}{1 + w^{\mathrm{T}} A^{-1} v} \tag{II,48}$$

где A — невырожденная $(n \times n)$-матрица; v, w — $(n \times 1)$-матрицы. Непосредственная проверка дает возможность с легкостью убедиться в правильности этой формулы. Полагая $v = y_j - B_j s_j$, $w = s_j$ и применяя формулу (II,48) к выражению (II, 47), после несложных преобразований получим:

$$H_{j+1} = H_j + \frac{(s_j - H_j y_j) s_j^{\mathrm{T}} H_j}{s_j^{\mathrm{T}} H_j y_j} \qquad (II,49)$$

Это известное преобразование Бройдена, которое из иных соображений было получено в работе [24]. Отметим ряд свойств метода Бройдена.

1-е свойство. В работе [29] показано, что метод Бройдена [см. выражения (II, 13), (II, 23), (II, 49)] позволяет находить решение системы линейных уравнений порядка n при числе шагов, не превышающем $2n$. При этом следует отметить, что если вместо формулы (II, 13) используется формула (II, 14), то метод Бройдена теряет это свойство.

2-е свойство. При использовании метода Бройдена для решения системы линейных уравнений (II,20) норма матрицы

$$C_{j+1} = A - B_{j+1} \qquad (II,50)$$

характеризующая ошибку аппроксимации матрицы A матрицей B_j, не возрастает с увеличением j. Действительно, подставим в уравнение (II, 47) значение y_j из условий (II, 35), а результат подставим в формулу (II, 50); тогда после несложных преобразований получим:

$$C_{j+1} = (A - B_j) \left[I_n - \frac{s_j s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} s_j} \right] \qquad (II,51)$$

Известно, что $(n - 1)$ собственное значение матрицы

$$I_n - \frac{s_j s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} s_j} \qquad (II,52)$$

равно единице, а одно равно нулю [24]. Матрица (II, 52) является симметрической, поэтому ее норма равна максимальному собственному значению [20, с. 330], т. е. единице. Но норма произведения двух матриц меньше или равна произведению их норм [20, с. 328]. Отсюда норма матрицы C_{j+1} меньше или равна норме матрицы C_j. Следовательно, с увеличением j норма матрицы C_j не возрастает. С другой стороны, доказательства, что $C_j \to 0$, а следовательно, B_j стремится к матрице A, нет. Отсюда следует, что для эффективной работы метода Бройдена требуется, чтобы в качестве B_0 использовалось хорошее приближение к матрице Якоби. Это объясняет, почему обычно в качестве B_0 используется разностная аппроксимация матрицы Якоби в начальной точке. Вообще говоря, в качестве B_0 может быть использована и единичная матрица, однако опыт показывает, что эффективность метода при этом существенно ухудшается [30].

3-е свойство. Если матрица H_i вырождена, то и все последующие матрицы H_k $(k > i)$ будут вырожденными. Докажем это свойство по индукции: пусть H_i вырождена; это значит, что ранг ее меньше n и существует ряд линейно независимых векторов q_j $(j = \overline{1, p})$, для которых справедливо равенство

$$H_i q_j = 0 \qquad (II,53)$$

Пусть Q — линейное подпространство, натянутое на векторы q_j $(j = \overline{1, p})$. Ясно, что если $q \in Q$, то

$$H_i q = 0 \qquad (II,54)$$

Умножив справа обе части равенства (II,49) на $q \in Q$ и используя (II, 54), получим

$$H_{i+1} q = 0 \qquad (II,55)$$

т. е. матрица H_{i+1} также является вырожденной. Отсюда следует, что начиная с точки i, все последующие направления поиска будут лежать в подпространстве $\overline{Q}$, ортогональном подпространству Q. Действительно, пусть $q \in Q$, тогда $p_i^{\mathrm{T}} q = -f_i^{\mathrm{T}} H_i q = 0$ и направление поиска p_i будет ортогонально любому вектору, принадлежащему Q. Аналогично доказывается, что и все последующие векторы p_k $(k > i)$ будут принадлежать подпространству $\overline{Q}$. Отсюда, между прочим, следует такой практический вывод. Если во время поиска матрица H_i по каким-либо причинам станет вырожденной, то все последующие направления поиска будут лежать в некотором подпространстве. Следовательно, и предельная точка также будет лежать в этом подпространстве. Поскольку маловероятно, чтобы истинное решение системы (II, 8) лежало в этом подпространстве, то в этом случае будет найдено неправильное решение. Это же свойство, по-видимому, объясняет следующее явление, наблюдающееся при использовании метода Бройдена. Часто случается, что последовательные приближения x_j изменяются крайне медленно. По-видимому, это связано с тем, что в процессе поиска матрицы H_i становятся близкими к вырожденным, и направления поиска лежат вблизи некоторого подпространства, хотя решение не принадлежит этому подпространству. Поэтому очень важно принимать меры, препятствующие вырождению матрицы H_i во время поиска.

В заключение характеристики метода Бройдена отметим, что описанные первые два его свойства, а также эффективное применение этого метода для решения разнообразных прикладных задач являются косвенным оправданием выбранного критерия (II, 41).

Заметим, что аппроксимация H_i может быть получена не через матрицу B_i, а непосредственно, в результате применения принципа наименьшего изменения матрицы H_{i-1}. Действительно, введем обозначение

$$H_{i+1} = H_i + D \qquad (II,56)$$

Тогда элементы матрицы D будем искать таким образом, чтобы ее норма Фробениуса была минимальна и при этом выполнялись ус-

ловия (II,31). Задача определения матрицы D запишется в этом случае следующим образом:

$$\min_{d_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2 \qquad (II,57)$$

$$Dy_j = m_j \qquad (II,58)$$

где $m_j = s_j - H_j y_j$. Действуя совершенно аналогично выводу формулы (II, 47), легко получить

$$H_{j+1} = H_j + \frac{(s_j - H_j y_j) y_j^{\mathrm{T}}}{y_j^{\mathrm{T}} y_i} \qquad (II,59)$$

Если сравнить формулы (II, 59) и (II, 47), то можно увидеть, что в них поменялись местами векторы s_j и y_j. Применение формулы (II,59) показало, что она работает неудовлетворительно [24]. В то же время практическое применение формулы Бройдена (II,49) для решения разнообразных прикладных задач показало ее большую эффективность.

Получим выражение для B_j при условии, что в качестве критерия минимизации будет использована норма Фробениуса некоторой взвешенной матрицы

$$V = \overline{W} E W \qquad (II,60)$$

где $\overline{W}$, W — произвольные невырожденные $(n \times n)$-матрицы. Итак, матрицу E ищем в данном случае как решение задачи

$$\min_{d_{ij}} \| \overline{W} E W \|_F \qquad (II,61)$$

$$Es_j = r_j \qquad (II,62)$$

формулы (II,60) имеем

$$E = \overline{W}^{-1} V W^{-1} \qquad (II,63)$$

Подставим в выражения (II,61), (II.62) значение E из формулы (II,63), при этом мы перейдем к новым поисковым переменным v_{ij}, т. е. к элементам матрицы V. После несложных выкладок придем к следующей задаче:

$$\min_{v_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} v_{ij}^2 \qquad (II,64)$$

$$Vc_j = \bar{r}_j \qquad (II,65)$$

где

$$\bar{r}_j = \overline{W} r_j \qquad c_j = W^{-1} s_j \qquad (II,66)$$

Заметим, что задача (II,64) выглядит так же, как задача (II,42), (II,43), поэтому по аналогии с выражением (II,46) получим

$$V = \frac{\bar{r}_j c_j^{\mathrm{T}}}{c_j^{\mathrm{T}} c_j} = \frac{\overline{W} (y_j - B_j s_j) (W^{-1} s_j)^{\mathrm{T}}}{(W^{-1} s_j)^{\mathrm{T}} (W^{-1} s_j)} \qquad (II,67)$$

Умножив полученное на $\overline{W}^{-1}$ слева и на W^{-1} справа, введя матрицу $M = (W^{-1})^{\mathrm{T}} W^{-1}$, вектор $v = Ms$ и используя формулу (II,38), получим

$$B_{j+1} = B_j + \frac{(y_j - B_j s_j) v^{\mathrm{T}}}{s_j^{\mathrm{T}} v} \qquad (II,68)$$

Используя формулу (II,48), получим выражение для матрицы H_{j+1}

$$H_{j+1} = H_j + \frac{(s_j - H_j y_j)\, v^{\text{т}} H_j}{v^{\text{т}} H_j y_j} \qquad (II,69)$$

Заметим, что, поскольку матрица M является произвольной и невырожденной, v есть произвольный вектор. Поэтому в формулах (II,68) и (II,69) его можно выбирать, что позволит улучшить сходимость метода. Прежде всего вектор v может быть выбран в соответствии с требованием максимальности знаменателя дроби в выражении (II,69), что будет препятствовать его обращению в нуль. В этом случае вектор v находится решением следующей экстремальной задачи:

$$\max_v v^{\text{т}} H_j y_j \qquad (II,70)$$

$$v^{\text{т}} v = 1 \qquad (II,71)$$

Поскольку второй член правой части равенства (II,69) является дробно рациональной функцией координат вектора v, величина этой дроби не зависит от абсолютной величины v. Для решения этой задачи воспользуемся методом множителей Лагранжа. Функция Лагранжа имеет вид:

$$L = v^{\text{т}} H_j y_j + 0{,}5\lambda\, (v^{\text{т}} v - 1) \qquad (II,72)$$

Приравняв производную L по v нулю, получим

$$v = -H_j y_j / \lambda \qquad (II,73)$$

Подставив это значение v в выражение (II,69), найдем

$$H_{j+1} = H_j + \frac{(s_j - H_j y_j)\,(H_j y_j)^{\text{т}} H_j}{(H_j y_j)^{\text{т}} H_j y_j}$$

Вектор v может быть выбран также из условия максимальности (по абсолютной величине) определителя матрицы B_{j+1}. Вынеся B_j в правой части формулы (II,68), легко получить

$$\det B_{j+1} = \det B_j \det \left[I_n + \frac{(B_j^{-1} y_j - s_j) v^{\text{т}}}{s_j^{\text{т}} v} \right] \qquad (II,74)$$

Воспользуемся следующей формулой [31]:

$$\det\,(I_n + ab^{\text{т}}) = 1 + b^{\text{т}} a \qquad (II,75)$$

где a, b — n-векторы. Применяя эту формулу, найдем

$$\det \left[I_n + \frac{(B_j^{-1} y_j - s_j)\, v^{\text{т}}}{s_j^{\text{т}} v} \right] = \frac{v^{\text{т}} B_j^{-1} y_j}{s_j^{\text{т}} v} \qquad (II,76)$$

Введем обозначения $d = B_j^{-1} y_j$; $b = s_j$, тогда задача выбора v сведется к следующей экстремальной задаче:

$$\max_v (\min) \frac{v^{\text{т}} d}{v^{\text{т}} b} \qquad v^{\text{т}} v = 1 \qquad (II,77)$$

Операцию максимизации необходимо проводить, если величина дроби окажется положительной и минимизацию — если отрицательной. Фактически, конечно, надо будет решать только одну из этих задач, поскольку метод множителей Лагранжа, дающий необходимые условия оптимальности, даст решение обеих задач.

Прямое применение метода множителей Лагранжа для решения задачи (II,77), приводит к необходимости решения кубических уравнений, что не очень удобно. Поэтому поступим следующим образом. Потребуем временно, чтобы знаменатель дроби в (II,77) был равным некоторой постоянной p. В этом случае задача (II,77) будет иметь вид:

$$\max_v \frac{v^{\text{т}} d}{p} \qquad (II,78)$$

$$v^{\text{т}} b = p \qquad v^{\text{т}} v = 1 \qquad (II,79)$$

Функция Лагранжа этой задачи записывается так:

$$L = v^{\text{т}} d/p + \lambda_1 v^{\text{т}} b + \lambda_2 v^{\text{т}} v$$

Приравняв производную $\partial L/\partial v$ нулю, подставляя полученное выражение для v в первую формулу (II,79) и исключая λ_2, получим квадратное уравнение для определения λ_1:

$$(d/p + \lambda_1 b)^{\text{т}} (d/p + \lambda_1 b) = (1/p^2) (d^{\text{т}} b/p + \lambda_1 b^{\text{т}} b)^2 \qquad \text{(II,80)}$$

Итак, процедура решения задачи (II,77) будет выглядеть следующим образом. Для фиксированного значения p из выражения (II,80) находятся два корня и соответствующие им два значения критерия $v^{\text{т}} d/p$. Из этих двух значений критерия выбирается наибольшее по абсолютной величине, которое и будет решением задачи (II,78), (II,79). Обозначим его через $(v^{\text{т}} d/p)^*$. Поскольку величина $(v^{\text{т}} d/p)^*$ является некоторой функцией параметра p, то для решения задачи (II,77) необходимо организовать поиск максимума величины $(v^{\text{т}} d/p)^*$ как функции одной переменной p.

Для определения вектора d надо решить систему линейных уравнений

$$B_j d = y_j \qquad \text{(II,81)}$$

Система уравнений (II,22) отличается от полученной только правой частью, но в любом случае она должна решаться. Поэтому решение системы (II,81) будет существенно более простой задачей.

Квазиньютоновские методы с постоянной глубиной памяти

Поскольку глубина памяти q в данном случае постоянна ($q > 1$), этот метод может начать работать после того, как каким-либо другим способом (скажем, квазиньютоновскими методами I-го ряда) накоплены векторы y_k, s_k ($k = \overline{0,\ q-1}$). Для определенности будем аппроксимировать саму матрицу Якоби. Матрицу B_{j+1} найдем в виде (II,38). В данном случае должны выполняться q соотношений (II,39). Для определения матрицы E применим принцип наименьшего изменения матрицы B_j задача определения матрицы E примет вид [26]:

$$\min_{e_{ij}} \sum_{k=1}^{n} \sum_{l=1}^{n} e_{kl}^2 \qquad \text{(II,82)}$$

$$E s_i = r_i \qquad i = j,\ j-1, \ldots,\ j-q+1 \qquad \text{(II,83)}$$

Функция Лагранжа имеет вид:

$$L = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} e_{kl}^2 + \sum_{p=1}^{q} \sum_{k=1}^{n} \lambda_k^{(p)} e_{kl} s_{l,\ j-p+1}$$

где $\lambda_k^{(p)}$ — k-тая компонента вектора множителей Лагранжа $\lambda^{(p)}$, соответствующего p-му условию (II,83). Приравнивая производные L по e_{ij} нулю, по аналогии с (II,45) легко получить

$$E = -\left(\lambda^{(1)} s_j + \lambda^{(2)} s_{j-1} + \cdots + \lambda^{(q)} s_{j-q+1}\right) \qquad \text{(II,84)}$$

Подставляя это значение E в q условий (II,83), получим уравнения для определения множителей Лагранжа:

$$\begin{array}{l} \lambda^{(1)} s_j^{\text{т}} s_j + \lambda^{(2)} s_{j-1}^{\text{т}} s_j + \cdots + \lambda^{(q)} s_{j-q+1}^{\text{т}} s_j = -r_j \\ \vdots \\ \lambda^{(1)} s_j^{\text{т}} s_{j-q+1} + \lambda^{(2)} s_{j-1}^{\text{т}} s_{j-q+1} + \cdots + \lambda^{(q)} s_{j-q+1}^{\text{т}} s_{j-q+1} = -r_{j-q+1} \end{array} \qquad \text{(II,85)}$$

Уравнения (II,85) являются системой из nq уравнений относительно nq неизвестных компонент векторов $\lambda^{(1)}, \ldots, \lambda^{(q)}$. Однако специальная структура этой системы поз-

воляет свести ее решеиие к существеиию более простой задаче. Действительно, будем временно считать величииы $\lambda^{(i)}$, r_i $(i = \overline{1, q})$ скалярами. Тогда система (II,85) превратится в систему из q уравиеиий с q иеизвестными. Обозиачим через Δ определитель системы (II,85). Тогда ее решеиие примет вид

$$\lambda^{(i)} = \sum_{k=1}^{n} \beta_{ki} r_{j-k+1}$$

где $\beta_{ki} = M_{ki}/\Delta$; M_{ki} — алгебраическое дополиеиие элемеита определителя Δ, стоящего на пересечеиии k-той строки и i-го столбца. Теперь снова будем считать $\lambda^{(i)}$ и r_i векторами. Простой подстаиовкой этих выражений для $\lambda^{(i)}$ в систему (II,85) легко убедиться в том, что эти формулы будут верны и в случае, когда $\lambda^{(i)}$ и r_i — векторы.

Квазиньютоновские методы 2-го рода

Линейные системы. Рассмотрим вначале случай, когда система (II, 8) является линейной [см. выражение (II,20)]. Приближения H_i к обратной матрице Якоби, т. е. в данном случае к матрице A^{-1}, будем находиться так, чтобы она удовлетворяла матричному уравнению (II,32) [25, 26]. Покажем, что на n-м шаге будет выполняться равенство

$$H_n = A^{-1} \quad \text{(II,86)}$$

если n построенных направлений $p_0, \ldots, p_{n-1}$ будут линейно независимыми. Действительно, пусть векторы p_j строятся с помощью формулы (II,23), в которой матрицы H_i удовлетворяют соотношению (II,32). На n-м шаге будет выполняться соотношение

$$H_n Y_n = S_n \quad \text{(II,87)}$$

При $i = n$ матрицы Y_i, S_i становятся квадратными. Ясно, что векторы s_j $(j = \overline{0, n-1})$ также являются линейно независимыми (они коллинеарны векторам p_j $(j = \overline{0, n-1})$, отсюда $\det S_n \neq 0$). Из уравнения (II,87) имеем $\det S_n = \det H_n \det Y_n$. Следовательно, $\det Y_n \neq 0$, $\det H_n \neq 0$ и существует обратная матрица Y_n^{-1}, а векторы $y_0, \ldots, y_{n-1}$ линейно независимы. Умножая равенство (II,87) на Y_n^{-1} и подставляя в полученное выражение значение S_n из (II,37), легко получить равенство (II, 86). Имеет место также следующий результат [32]: если $\{p_j\}_0^{(i-1)}$ — семейство векторов $p_0, \ldots, p_{i-1}$ и H_i — неособенная матрица, удовлетворяющая условию (II,32), то для линейной системы (II,20) вектор p_i, построенный с помощью формулы (II,32), либо является линейно независимым относительно $\{p_j\}_0^{(i-1)}$, либо точка x_{i+1} является решением системы (II,20).

Рассмотрим теперь вопрос об определении матрицы H_i. Введем две вспомогательные матрицы K_i^j $(j = 1, 2)$ следующим образом

$$K_i^j = I_n - Y_i Y_i^j \quad \text{(II,88)}$$

где Y_i^j $(j = 1, 2)$ g-обратные матрицы [11, с. 266], удовлетворяющие соотношениям

$$Y_i Y_i^j Y_i = Y_i \quad \text{(II,89)}$$

Воспользовавшись формулой общего решения системы уравнений (II,32) [11, с. 267] и подходом, развитым в работе [33] (см. также [11, с. 66—70]), можно получить следующие рекуррентные формулы для определения матрицы H_i:

$$H_{i+1} = H_i + \frac{s_i d_{1i}^{\mathrm{T}} K_i^1}{d_{1i}^{\mathrm{T}} K_i^1 y_i} - \frac{\left(H_i - RK_i^2\right) y_i c_{1i}^{\mathrm{T}} K_i^1}{c_{1i}^{\mathrm{T}} K_i^1 y_i} -$$

$$- R\left[\frac{y_i d_{2i}^{\mathrm{T}} K_i^2}{d_{2i}^{\mathrm{T}} K_i^2 y_i} - \frac{\left(I_n - K_i^2\right) y_i c_{2i}^{\mathrm{T}} K_i^2}{c_{2i}^{\mathrm{T}} K_i^2 y_i}\right] \qquad (II,90)$$

$$K_{i+1}^j = K_i^j - \frac{\left(K_i^j - I_n\right) y_i c_{ji}^{\mathrm{T}} K_i^j}{c_{ji}^{\mathrm{T}} K_i^j y_i} - \frac{y_i d_{ji}^{\mathrm{T}} K_i^j}{d_{ji}^{\mathrm{T}} K_i^j y_i} \qquad (II,91)$$

где R_i — произвольная $(n \times n)$-матрица.

$$K_0^j = I_n \qquad (II,92)$$

$$H_0 = R \qquad (II,93)$$

Матрицы K_i^j обладают следующими свойствами.

Если направления $s_0, \ldots, s_{n-1}$ линейно независимы, то выполняются равенства

$$1.\ K_n^j = 0 \qquad j = 1,\ 2 \qquad (II,94)$$

$$2.\ K_i^j K_i^j = K_i^j \qquad j = 1,\ 2 \qquad (II,95)$$

$$3.\ K_i^j\left(I_n - K_i^j\right) = 0 \qquad j = 1,\ 2 \qquad (II,96)$$

Равенства (II,95), (II,96) просто доказываются подстановкой в них в величины K_i^j из уравнения (II, 88) и применением равенства (II,89).

Остановимся подробнее на первом свойстве. Как было показано выше, из линейной независимости векторов $s_0, \ldots, s_{n-1}$ следует существование обратной матрицы Y_n^{-1}. Тогда, из соотношения (II,89) при $i = n$ получим $Y_n Y_n^j = I_n$. Откуда, учитывая формулу (II,88), получим равенство (II, 94).

Построение частных формул определения матрицы H_i. Соотношения (II,90), (II,91) являются достаточно общими формулами построения матрицы H_i, зависящими от произвольной матрицы R и произвольных векторов c_{ji}, d_{ji} $(j = 1, 2)$. Давая произвольным величинам определенные значения, можно получить частные формулы для определения H_i. Совокупность этих формул обычно называют семейством формул Адачи [33]. Если положить $c_{1i} = d_{1i}$ и $c_{2i} = d_{2i}$ в формулах (II,90), то

$$H_{i+1} = H_i + \frac{\left(s_i - H_i y_i + RK_i^2 y_i\right) c_{1i}^{\mathrm{T}} K_i^1}{c_{1i}^{\mathrm{T}} K_i^1 y_i} - \frac{RK_i^2 y_i c_{2i}^{\mathrm{T}} K_i^2}{c_{2i}^{\mathrm{T}} K_i^2 y_i} \qquad (II,97)$$

$$K_{i+1}^j = K_i^j - \frac{K_i^j y_i c_{ji}^{\mathrm{T}} K_i^j}{c_{ji}^{\mathrm{T}} K_i^j y_i} \qquad j = 1,\ 2 \qquad (II,98)$$

Если

$$c_{1i} = c_{2i} = c_i \qquad d_{1i} = d_{2i} = d_i$$

то

$$K_i^1 = K_i^2 = K_i$$

$$H_{i+1} = H_i + \frac{(s_i - Ry_i)\, d_i^{\text{т}} K_i}{d_i^{\text{т}} K_i y_i} - \frac{(H_i - R)\, y_i c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i y_i} \tag{II,99}$$

$$K_{i+1} = K_i - \frac{(K_i - I_n)\, y_i c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i y} - \frac{y_i d_i^{\text{т}} K_i}{c_i^{\text{т}} K_i y_i} \tag{II,100}$$

Если в выражении (II, 100) $c_i = d_i$, то

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i y_i} \tag{II, 101}$$

$$K_{i+1} = K_i - \frac{K_i y_i c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i y_i} \tag{II, 102}$$

Заметим, что поскольку существует равенство (II, 86), то любая формула из семейства (II, 90), (II, 91) может быть использована для определения обратной матрицы, а также для решения систем линейных уравнений. Эти формулы особенно полезны в том случае, когда явный вид матрицы A и вектора b в системе линейных уравнений (II, 20) нам неизвестен, и мы можем найти $f(\bar{x}) = A\bar{x} + b$ только при заданном $\bar{x}$. Такая ситуация может иметь место, когда модели блоков ХТС линейны и используется последовательный метод расчета ХТС. Действительно, в этом случае система уравнений относительно итерируемых переменных (II, 5) будет иметь вид (II, 20), в котором явный вид $(n \times n)$-матрицы A и вектора b нам неизвестен, и мы можем найти $f(\bar{x}) = A\bar{x} + b$ только по заданному $\bar{x}$, зная модели блоков и последовательность их расчета. Используя любую из формул семейства (II, 90), (II, 91) совместно с уравнениями (II, 14), (II, 23), мы на n-м шаге получим матрицу A^{-1} и решение системы (II, 20) (см. с. 41).

Так же, как были выведены формулы (II, 90)—(II, 102) для аппроксимации матрицы обратной матрице Якоби, могут быть выведены формулы для аппроксимации B_i самой матрицы Якоби. Для этого к системе (II, 29) надо применить общую формулу решения матричных уравнений [11, с. 267]. Рассуждая так же, как и в предыдущем случае, легко получить аналоги формул (II, 90)—(II, 102) для матрицы B_i. Легко видеть, что формально эти выражения могут быть получены из формул (II, 90)—(II, 102), если поменять местами величины s_i и y_i. Выпишем здесь только аналоги формул (II, 101), (II, 102):

$$B_{i+1} = B_i + \frac{(y_i - B_i s_i)\; c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i s_i} \tag{II, 103}$$

$$K_{i+1} = K_i - \frac{K_i s_i c_i^{\text{т}} K_i}{c_i^{\text{т}} K_i s_i} \tag{II, 104}$$

Так же, как было доказано равенство (II, 86), можно показать, что при $i = n$ будет выполняться равенство

$$B_n = A \tag{II, 105}$$

Остановимся подробнее на применении формул (II, 101), (II, 102) или (II, 103), (II, 104). В них имеется произвольный вектор c_i. Единственное условие, которому должен удовлетворять этот вектор, состоит в том, чтобы знаменатель в выражениях (II, 101), (II, 102) был отличен от нуля. В работе [33] в качестве c_i рекомендуется поочередно выбирать столбцы единичной матрицы I_n. Однако более правильно выбирать c_i, чтобы улучшить сходимость и предотвратить появление нежелательных явлений. Для выбора c_i здесь могут быть привлечены те же самые соображения, что и при выборе вектора v в формуле (II, 70). Можно выбирать c_i так, чтобы знаменатель в выражениях (II, 101), (II, 102) был максимальным, что будет препятствовать его стремлению к нулю. В этом случае определение вектора c_i будет подобно решению задачи (II, 71), (II, 72). По аналогии с формулой (II, 73) ее решение будет иметь вид $c_i = -K_i y_i/\lambda$ (λ — множитель Лагранжа). Подставляя это значение в выражения (II, 101), (II, 102), найдем:

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, y_i^{\mathrm{T}} K_i^{\mathrm{T}} K_i}{y_i^{\mathrm{T}} K_i^{\mathrm{T}} K_i y_i} \qquad K_{i+1} = K_i - \frac{K_i y_i y_i^{\mathrm{T}} K_i^{\mathrm{T}} K_i}{y_i^{\mathrm{T}} K_i^{\mathrm{T}} K_i y_i} \tag{II, 106}$$

Непосредственной проверкой легко убедиться, что если $K_i^{\mathrm{T}} = K_i$, то и $K_{i+1}^{\mathrm{T}} = K_{i+1}$, т. е. если $K_0^{\mathrm{T}} = K_0$, то и все последующие матрицы K_i будут симметричными. Используя симметричность K_i и равенство (II, 95), преобразуем соотношения (II, 106)

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, y_i^{\mathrm{T}} K_i}{y_i^{\mathrm{T}} K_i y_i} \tag{II, 107}$$

$$K_{i+1} = K_i - \frac{K_i y_i y_i^{\mathrm{T}} K_i}{y_i^{\mathrm{T}} K_i y_i} \tag{II, 108}$$

Введем обозначение

$$a_i = K_i y_i \tag{II, 109}$$

и перепишем формулу (II, 107) в виде

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, a_i^{\mathrm{T}}}{a_i^{\mathrm{T}} y_i} \tag{II, 110}$$

В главе III будет показано, что матрица K_i, имеющая структуру (II, 108) [см. выражение (III, 54)], является оператором проектирования на подпространство C, ортогональное подпространству, натянутому на векторы $y_0, \ldots, y_{i-1}$, а вектор a_i является ортогональной проекцией y_i на подпространство C, т. е.

$$a_i^{\mathrm{T}} Y_i = 0 \tag{II, 111}$$

Верно и обратное, т. е. если при всех i вектор a_i удовлетворяет условию (II, 111), то матрица H_i будет удовлетворять уравне-

нию (II, 32). Действительно, пусть на i-том шаге выполняется соотношение (II, 32) и вектор a_i удовлетворяет условию (II, 111), тогда и H_{i+1} из соотношения (II, 110) будет удовлетворять уравнению $H_{i+1}Y_{i+1} = S_{i+1}$ или

$$H_{i+1}y_j = s_j \qquad j = i,\ i-1, \ldots, 0 \qquad \text{(II, 112)}$$

Действительно, умножив (II, 110) на y_i, получим:

$$H_{i+1}y_i = H_i y_i + (s_i - H_i y_i) = s_i$$

Следовательно, соотношение (II, 112) для $j = i$ доказано. Умножим теперь уравнение (II, 110) на y_k ($k \leqslant i-1$), и в соответствии с выражениями (II, 111), (II, 32) получим $H_{i+1}y_k = H_i y_k = s_k$ ($k \leqslant i-1$), что подтверждает справедливость соотношения (II, 112) при $j < i$. Проведя аналогичные рассуждения для случая, когда ищется приближение к самой матрице Якоби, получим формулу, аналогичную соотношению (II, 110) для определения матриц B_i

$$B_{i+1} = B_i + \frac{(y_i - B_i s_i)\, b_i^{\mathrm{T}}}{b_i^{\mathrm{T}} s_i} \qquad \text{(II, 113)}$$

где b должно удовлетворять условию $b_i^{\mathrm{T}} S_i = 0$.

В случае, когда вектор b_i будет выбираться из условия максимальности знаменателя в правой части формулы (II, 113), он (вектор) должен определяться по формуле

$$b_i = P_i s_i \qquad P_{i+1} = P_i - \frac{P_i s_i s_i^{\mathrm{T}} P_i}{s_i^{\mathrm{T}} P_i s_i} \qquad \text{(II, 114)}$$

Если строится аппроксимация к самой матрице Якоби [см. формулы (II, 103), (II, 104)] очень важно, чтобы матрицы B_i не становились вырожденными. В этом случае может оказаться целесообразным выбирать c_i из условия максимальности $\det B_{j+1}$ по абсолютной величине, а способ определения c_1 будет очень похож на способ определения вектора v, обеспечивающий максимум абсолютной величины детерминанта матрицы B_{j+1} в выражении (II, 69). Поэтому на данном вопросе мы остановимся очень кратко. Вынеся матрицу B_i в правой части равенства (II, 113) за скобку, и воспользовавшись правилом вычисления определителя произведения матриц и формулой (II, 75), придем к задаче (II, 77), в которой

$$d = K_i B_i^{-1} y_i \qquad v = c_i \qquad b = K_i y_i$$

Отметим еще один очень возможный способ выбора параметров c_i. Как уже указывалось (см. с. 37) полезным оказался принцип наименьшего изменения аппроксимирующей матрицы, на основе которого были получены выражения для матриц E и D в квазиньютоновских методах 1-го рода. В данном случае вид матриц E и D известен. Его легко получить из формул (II, 101), (II, 103). Произвольные константы c в этих формулах можно попытаться определить, исходя из этого принципа.

Нелинейные системы. Решение нелинейной системы (II, 8), в отличие от линейной, не будет получено за n шагов. Поэтому итерационный процесс, определяемый формулами (II, 14), (II, 23), (II, 90) и (II, 91) должен быть продолжен при $i > n$. Однако в соответствии с равенством (II, 94) матрица K_n^j становится нулевой при $i = n$ и формула (II, 91) теряет смысл. Здесь можно поступить следующим образом. Мы можем считать, что в результате n шагов закончен некоторый цикл и мы получили некоторое приближение H_n к матрице J_n^{-1}; далее в качестве начального значения H_0 используется полученное значение H_n, а матрицы K_i^j берутся равными единичным. В конце этого цикла опять проделывается та же процедура, и т. д. Чтобы выразить математически эту процедуру, для матриц H_i, K_i^j удобно ввести второй индекс l (номер цикла) и записывать их в виде $H_{i,l}$, $K_{i,l}^j$. Тогда внутри каждого цикла индекс i (номер шага) будет меняться от 1 до n, а индекс l будет изменяться на единицу через каждые n шагов. Тогда итерационный процесс, определяемый формулами (II, 14), (II, 23), (II, 90), (II, 91) можно представить состоящим из циклов. Внутри цикла матрицы $H_{i,l}$, $K_{i,l}^j$ преобразуются в соответствии с формулами (II, 90), (II, 91), а в начальной точке каждого цикла матрицы $H_{i,l}$, $K_{i,l}^j$ определяются следующим образом:

$$H_{n,\, l-1} = H_{0,\, l} \qquad K_{0,\, l}^i = I_n \tag{II, 115}$$

Рассмотрим применение этого циклического итерационного процесса в случае линейных уравнений (II, 20). Как было показано, применение любой из формул семейства (II, 90), (II, 91) приведет к тому, что при $i = n$ будет выполняться равенство (II, 86). В соответствии с циклическим способом применения формул (II, 90), (II, 91) при $i = kn$ $(k = 1, 2, \ldots)$ будут выполняться равенства $H_{k,n} = A^{-1}$. Рассмотрим теперь использование формул (II, 103), (II, 104). Применяя формулу (II, 103) при $i = n$ и используя выражение (II, 35) и равенство $B_n = A$, легко найти, что $B_{n+1} = A$. Аналогичным образом можно показать, что при всех $i > n$ будет выполняться равенство

$$B_i = A \qquad i > n \tag{II, 116}$$

Расчет стационарных режимов химико-технологического процесса Вильямса—Отто

В этом и следующих разделах приведены результаты численных расчетов на ЭВМ стационарных режимов ряда химико-технологических процессов. При проведении расчетов использовались, в основном, квазиньютоновские методы решения систем нелинейных уравнений: *QNM*, определяемый преобразованием (II, 101) и *Broyden* [см. преобразование (II, 49)]. В соответствующих алгоритмах каждая следующая итерационная точка в пространстве независимых

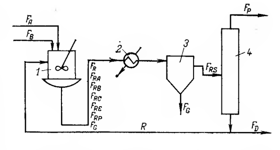


Рис. 7. Схема процесса Вильямса — Отто:
1 — реактор идеального перемешивания; *2* — теплообменник; *3* — фильтр; *4* — ректификационная колонна.

переменных определялась условием, чтобы вычисленное в этой точке значение нормы вектора левых частей системы уравнений было меньше, чем в предыдущей. В качестве начального приближения H_0 к обратной матрице Якоби выбиралась либо разностная аппроксимация матрицы Якоби в начальной точке с последующим обращением (соответствующие варианты методов обозначаются далее *QNM1*, *Broyden 1*), либо единичная матрица I_n (соответствующие варианты обозначены через *QNM2*, *Broyden2*). Некоторые из приводимых результатов получены методом простой итерации (в случае сходимости). Применялись также и другие методы решения систем нелинейных уравнений: *DEM* [37], *GDEM* [23], метод Вольфа [3, с. 35].

Рассмотрим, ставший уже традиционным, модельный химико-технологический процесс Вильямса—Отто [38; 39]; получаемый продукт обозначаем через *P* (рис. 7). Схема состоит из реактора идеального перемешивания *1*, теплообменника *2*, фильтра *3* и ректификационной колонны *4*. В реакторе протекают следующие необратимые реакции второго порядка:

$$A + B \xrightarrow{k_1} C \qquad C + B \xrightarrow{k_2} P + E \qquad P + C \xrightarrow{k_3} G \quad \text{(II, 117)}$$

где k_i — константы скоростей реакций. Реагенты A, B — сырьевые продукты. Поток из реактора *1* подается в теплообменник *2*, где охлаждается до температуры, достаточной для полного отделения побочного продукта G в фильтре *3*. Поток с остальными компонентами подается на вход ректификационной колонны *4*, из верхней части которой отбирается продукт P. Известно, что компоненты P и E образуют азеотропную смесь, в которой относительная доля P составляет 10 % (масс.). Это определяет режим работы колонны и учитывается при ее моделировании. Поток, выходящий из нижней части колонны *4*, разделяется на две части: одна выводится из схемы и утилизируется, вторая возвращается в цикл, на вход реактора *1*.

Задачу расчета режима данной замкнутой ХТС при заданных расходах сырья F_A, F_B, объеме реактора ($V = 60$ условных единиц), температуре в реакторе T и коэффициенте рецикла α рассмотрим, как задачу решения системы нелинейных уравнений относительно величин F_{RA}, F_{RB}, F_{RC}, F_{RE}, F_{RP}, выражающих концентрации компонент A, B, C, E, P в выходном потоке реактора. Соот-

ветствующая система уравнений формируется на основе математических моделей элементов схемы:

для реактора

$$F_{RA} = F_A + R_A - \frac{k_1 F_{RA} F_{RB} V \rho}{F_R^2} \quad (II, 118)$$

$$F_{RB} = F_B + R_B + \frac{(-k_1 F_{RA} F_{RB} - k_2 F_{RB} F_{RC}) V \rho}{F_R^2} \quad (II, 119)$$

$$F_{RC} = R_C + \frac{(2k_1 F_{RA} F_{RB} - 2k_2 F_{RB} F_{RC} - k_3 F_{RC} F_{RP}) V \rho}{F_R^2} \quad (II, 120)$$

$$F_{RE} = R_E + \frac{2k_2 F_{RB} F_{RC} V \rho}{F_R^2} \quad (II, 121)$$

$$F_{RP} = R_P + \frac{(k_2 F_{RB} F_{RC} - 0{,}5 k_3 F_{RC} F_{RP}) V \rho}{F_R^2} \quad (II, 122)$$

где

$$F_R = F_A + F_B + R_A + R_B + R_C + R_E + R_P \quad (II, 123)$$

ρ — плотность (ρ = 50 условных единиц); k_i — константы скоростей реакций, рассчитываемые по формулам $k_i = A_i^* \exp(-B_i^*/T)$; T — температура в реакторе (значения коэффициентов A_i^* и B_i^* даны в табл. 2).

для ректификационной колонны

$$F_{si} = F_{Ri} \qquad i = A, B, C, E \qquad F_{sP} = 0{,}1 F_{RE} \quad (II, 124)$$

для делителя потока (расположен после колонны)

$$R_i = \alpha F_{si} \qquad i = A, B, C, E, P \quad (II, 125)$$

В связи с тем, что при решении этой системы метод простой итерации дает расходящуюся последовательность приближений, рассматривалась также и видоизмененная система уравнений для реактора

$$F_{RA} = \frac{F_A + R_A}{1 + k_1 F_{RB} V \rho F_R^{-2}} \qquad F_{RB} = \frac{F_B + R_B}{1 + (k_1 F_{RA} + k_2 F_{RC}) V \rho F_R^{-2}}$$

$$F_{RC} = \frac{R_C + 2k_1 F_{RA} F_{RB} V \rho F_R^{-2}}{1 + (2k_2 F_{RB} + k_3 F_{RP}) V \rho F_R^{-2}} \qquad F_{RE} = R_E + 2k_2 F_{RB} F_{RC} V \rho F_R^{-2}$$

$$F_{RP} = \frac{R_P + k_2 F_{RB} F_{RC} V \rho F_R^{-2}}{1 + 0{,}5 k_3 F_{RC} V \rho F_R^{-2}} \quad (II, 126)$$

В этом случае метод простой итерации обеспечивал сходимость. При всех методах расчет начинался из 6 начальных точек. Значения, принимаемые параметрами F_A, F_B, α, T и итерируемыми переменными F_{Ri} ($i = A, B, C, E, P$) в этих точках, приведены в табл. 3. Данные по расчету схемы приведены: в случае использования модели (II, 118)—(II, 123) — в таблице 4, а в случае использова-

ния модели (II,126) — в таблице 5. В первой строке таблицы 4 (а также таблицы 5) даны значения нормы вектора невязок системы уравнений модели в начальной точке — $NORM_0$, в остальных — число обращений к расчету модели.

Таблица 2. **Параметры констант скоростей**

i	A_i^*	B_i^*
1	$5{,}9755 \cdot 10^9$	$12 \cdot 10^3$
2	$2{,}5962 \cdot 10^{12}$	$15 \cdot 10^3$
3	$9{,}6283 \cdot 10^{15}$	$20 \cdot 10^3$

Чтобы получить начальное приближение H_0 для обратной матрицы Якоби системы (II,118) — (II,123) или (II,126), производные левых частей систем нелинейных уравнений аппроксимировались с помощью конечных разностей (с последующим обращением полученной матрицы). В соответствии с рекомендациями [24] приращения независимых переменных Δx_i выбирались равными: $\Delta x_i = 0{,}001 x_{0,i}$ (здесь $x_{0,i}$ — i-тая координата начальной точки x_0). При этом затраты на вычисление разностной аппроксимации матрицы Якоби в начальной точке эквивалентны $(n+1)$-му расчету левых частей системы уравнений.

При использовании хорошего начального приближения для x_0 (см. табл. 3, столбец для начальной точки 3) методы QNM1 и QNM2 давали такие же (либо лучшие) результаты, что и соответствующие

Таблица 3. **Значения переменных в начальных точках**

Переменная	Значения переменных для начальных точек 1—6					
	1	2	3	4	5	6
F_A	10 000	11 540	13 546	13 546	11 540	10 000
F_B	40 000	31 230	31 523	31 523	31 230	40 000
α	0,74	0,58	0,75	0,896	0,896	0,896
F_{RA}	6 000	8 820	18 187	38 420	38 493	38 251
F_{RB}	35 000	39 910	60 815	121 745	121 833	120 818
F_{RC}	10 000	2 360	3 331	7 296	7 311	7 266
F_{RE}	15 000	31 660	60 542	137 541	137 708	136 416
F_{RP}	6 000	7 890	10 817	18 517	18 534	18 405
T	610	610	656	654	654	654

Таблица 4. **Результаты расчета схемы Вильямса—Отто по уравнениям (II,118)—(II,125)**

Метод	Результаты для начальных точек 1—6					
	1	2	3	4	5	6
$NORM_0$	29 286	16 103	428	11 302	17 735	85 502
QNM1	14	14	10	26	15	23
Broyden 1	15	14	10	25	20	25
QNM2	21	21	23	28	27	34
Broyden 2	55	41	49	89	94	136

Таблица 5. **Результаты расчета схемы Вильямса—Отто по уравнениям (II,124)—(II,126)**

Метод	Результаты для начальных точек 1—6					
	1	2	3	4	5	6
$NORM_0$	23 159	15 085	331	9 661	14 234	59 010
QNM1	16	14	11	112	80	29
Broyden 1	17	14	11	92	155	31
QNM2	23	21	19	25	22	37
Broyden 2	88	52	53	170	144	117
Простая итерация	105	66	114	281	285	283

варианты метода Бройдена (см. табл. 4 и табл. 5, столбцы для начальных точек 3). При использовании плохого начального приближения для (см. табл. 3, столбцы для начальных точек 1—2, 4—6) метод QNM во многих случаях давал существенно лучшие результаты, чем метод Бройдена. Расчет схемы (см. рис. 7) с помощью метода QNM2 показал, что число вычислений незначительно увеличивается при переходе от хорошего приближения к плохому.

Интересно отметить, что методы Бройдена и QNM сходились, когда метод простой итерации расходился, и давали существенно лучшие результаты, когда метод простой итерации сходился. Причем применение метода Бройдена и QNM для расчета варианта математической модели схемы, приведенной на рис. 7, для которого процесс простой итерации расходился, давало лучшие результаты, чем применение этих же методов для расчета варианта математической модели, для которого простая итерация обеспечивала сходимость.

*Расчет стационарных режимов химико-технологической системы изомеризации н-пентана**

Краткое описание технологического процесса. ХТС изомеризации *н*-пентана предназначена для получения изопентана высокотемпературным способом [40, с. 85]. Целевой продукт (изопентан) является остродефицитным, вследствие его широкого использования в качестве растворителя (производства изопренового каучука и бутилкаучука) в качестве компонента высокооктановых бензинов и для других целей. Технологический процесс производства изопентана представляет собой замкнутую химико-технологическую схему с материальными и тепловыми рециклами, что обусловлено современными требованиями рекуперации тепла и использования непрореагировавшего сырья; схема состоит из следующих основных узлов: азеотропная осушка исходной *н*-пентановой фракции, изомеризация *н*-пентана, водородсодержащего газа (ВСГ), комприми-

* Раздел написан совместно с В. Б. Покровским и Н. Н. Зиятдиновым.

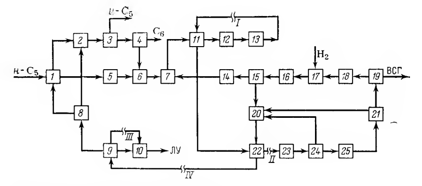


Рис. 8. Схема ХТС изомеризации *н*-пентана:

1, 9, 11, 22 — рекуперативные теплообменники; *2, 6, 7, 17, 20* — смесители потоков; *3, 4, 10* — ректификационные колонны; *5* — колонна азеотропной осушки; *8, 19* — делители потоков; *12* — пароперегревательная печь; *13* — реактор; *14* — адсорбер; *15, 21, 24* — сепараторы; *16, 23, 25* — теплообменники; *18* — компрессор; *I—IV* — номера разрываемых потоков.

н-C_5 — сырье (*н*-пентановая фракция); *и*-C_5 — продукт (изопентановая фракция); C_6 — гексановая фракция; ЛУ — легкие углеводороды; H_2 — свежий водород; ВСГ — водородсодержащий газ.

рование и осушка ВСГ, ректификация продуктов реакции (рис. 8). Сырье, *н*-пентановая фракция, после колонны азеотропной осушки *5* смешивается с *н*-пентановой фракцией рецикла и циркулирующим ВСГ, нагревается в рекуператоре *11* и печи *12* и поступает на изомеризацию в реактор *13*. Процесс изомеризации протекает при температуре 360—450 °С и давлении $35 \cdot 10^6$ Па в реакторе с неподвижным слоем алюмоплатинового катализатора в среде ВСГ. Побочной реакцией является крекинг *н*-пентана и изопентана в легкие углеводороды. Контактный газ из реактора охлаждается и конденсируется в рекуператорах *11*, *22*, холодильниках *23*, *25*, отделяется от ВСГ в сепараторах *24*, *21* и через рекуператоры *22*, *9* поступает в ректификационную колонну *10*, где происходит выделение легких углеводородов (ЛУ). Кубовый продукт колонны *10*, изопентан-пентановая фракция, через рекуператоры *9*, *1* поступает в ректификационную колонну *3* на выделение целевого продукта (изопентан). Кубовый продукт колонны *3* (непрореагировавший в реакторе *н*-пентан) поступает в ректификационную колонну *4* на выделение фракции гексан и выше, и далее в качестве рецикла смешивается со свежим *н*-пентаном. Циркулирующий ВСГ поступает на всасывание компрессора *18*. На выходе компрессора сжатый ВСГ смешивается со свежим водородом и, пройдя холодильник *16*, сепаратор *15* и адсорбер *14*, поступает на смешение с *н*-пентановой фракцией.

Математические модели аппаратов приведены ниже.

Реактор 13. Реакция высокотемпературной изомеризации *н*-пентана на бифункциональном катализаторе протекает через стадии дегидрирования *н*-пентана в олефин, изомеризации олефина в изоолефин, гидрирования изоолефина в изопентан. Учитывая, что лимитирующей стадией реакции является изомеризация олефина [41] и что, кроме основной реакции, процесс сопровождается необра-

тимым гидрокрекингом $н$-пентана и изопентана в легкие углеводороды (метан-бутановая фракция), упрощенный механизм реакции может быть представлен в виде

$$X_1 \underset{w_2}{\overset{w_1}{\rightleftarrows}} X_2 \qquad X_1 \xrightarrow{w_3} \alpha_1 X_3 \qquad X_2 \xrightarrow{w_4} \alpha_2 X_3 \qquad (II,\ 127)$$

где X_1 — $н$-пентан; X_2 — изопентан; X_3 — легкие углеводороды; w_i — скорость i-той реакции; α_1, α_2 — стехиометрические коэффициенты. Реакции крекинга играют незначительную роль вследствие того, что реакция изомеризации протекает с малым тепловым эффектом и высокой селективностью. Этим обусловливается низкий суммарный тепловой эффект процесса, составляющий 84 кДж/кг сырья. Влияние параметров процесса на изомеризацию $н$-пентана описано в [42]. Скорости реакции изомеризации и побочных реакций описываются уравнениями

$$w_1 = k_1 p_1/p_{H_2} \qquad w_2 = k_2 p_2/p_{H_2} \qquad w_3 = k_3 p_1 \qquad w_4 = k_4 p_2 \quad (II,\ 128)$$

где p_1, p_2, p_{H_2} — парциальные давления $н$-пентана, изопентана и водорода соответственно; k_i $(i = 1,4)$ — кажущиеся константы скоростей реакций, зависимость которых от температуры выражается уравнением Аррениуса: $k_i = k_{0i} \exp(-E_i/RT)$, в котором R — газовая постоянная; T — абсолютная температура; величины k_{0i} [кмоль/(ч·м^3·МПа)], R [кДж/(моль·К)] и E_i [кДж/моль] равны

$$k_{02} = 0{,}92 \cdot 10^{11} \qquad k_{03} = 0{,}12 \cdot 10^{18} \qquad k_{04} = 0{,}45 \cdot 10^{16}$$

$$E_2 = 139{,}0 \qquad E_3 = 330{,}8 \qquad E_4 = 221{,}9 \qquad R = 8{,}3 \cdot 10^{-3}$$

Константа равновесия реакции изомеризации $k = k_1/k_2$ в рабочей области температур аппроксимируется уравнением [40, с. 72]: $k = 2{,}89 + 0{,}0044\,(600 - T)$.

В математическом описании реактора изомеризации $н$-пентана приняты следующие допущения: гидродинамическая обстановка в промышленном аппарате близка к потоку идеального вытеснения; тепловой режим является адиабатическим; активность катализатора стабильна в течение длительного времени, тепловым балансом можно пренебречь. В окончательном виде математическое описание, полученное интегрированием исходной системы дифференциальных уравнений, выглядит так:

$$x_1 = c_1\,[r_1 + A\,(k_2 + k_4 p)] \exp(r_1\varphi) + c_2\,[r_2 + A\,(k_2 + k_4 p)] \exp(r_2\varphi) \qquad (II,\ 129)$$

$$x_2 = Akk_2\,[c_1 \exp(r_1\varphi) + c_2 \exp(r_2\varphi)] \qquad t = t^0 + \frac{q\,(x_1^0 - x_1)}{c} \qquad (II,\ 130)$$

$$x_3 = \frac{Ap\,\{\alpha_1 k_3\,[r_1 + A\,(k_2 + k_4 p)] + \alpha_2 k_4 Akk_2\}\,[\exp(r_1\varphi) - 1]\,c_1}{r_1} +$$

$$+ \frac{Ap\,\{\alpha_1 k_3\,[r_2 + A\,(k_2 + k_4 p)] + \alpha_2 k_4 A k_2 k\}\,[\exp(r_2\varphi) - 1]\,c_2}{r_2} + x_3^0 \qquad (II,\ 131)$$

$$D = \sqrt{[k_2\,(k - 1) + (k_3 - k_4)\,p]^2 + 4kk_2^2} \qquad (II,132)$$

$$r_{1,2} = \{-A\,[k_2\,(k + 1) + (k_3 + k_4)\,p] \pm AD\}/2 \qquad A = \rho_к V_к RT/Gp \quad (II,\ 133)$$

$$c_1 = \frac{Akk_2 x_1^0 - x_2^0\,[r_2 + A\,(k_2 + k_4 p)]}{A^2 kk_2 D} \qquad c_2 = -c_1 + \frac{x_2^0}{Akk_2} \qquad (II,\ 134)$$

где $\varphi = 1$; x_i^0 $(i = 1, 3)$ — содержание *н*-пентана, изопентана, легких углеводородов на входе в реактор, мольн. доли; $\rho_к$ — плотность катализатора, кг/м³; $V_к$ — объем катализатора, м³; G — расход смеси в реактор, кг/ч; $p \approx p_{H_2}$ — парциальное давление водорода в реакторе, МПа; t^0, t — температуры на входе в реактор и выходе из него, °С; q — суммарный тепловой эффект, кДж/кмоль; c — теплоемкость *н*-пентана, кДж/(кмоль °С).

Ректификационные колонны 3, 4, 10. В математической модели ректификационных колонн приняты следующие допущения: разделяемая смесь — бинарная; жидкость в колонне находится при температуре кипения, а пар насыщенный; конденсатор для колонн *3*, *4* — полный, для колонны *10* — парциальный; колонна работает в адиабатических условиях; массопередача на тарелках эквимолярная; к. п. д. тарелок принимается постоянным по всей колонне; относительная летучесть смеси постоянна по высоте колонны. Математическая модель ректификационной колонны представляет собой следующую систему уравнений [43, с. 18].

Уравнение общего материального баланса колонны

$$F = D + W \qquad \text{(II, 135)}$$

Уравнение материального баланса колонны по легколетучему компоненту

$$Fx_F = Dx_D + Wx_W \qquad \text{(II, 136)}$$

Уравнение материального баланса i-той тарелки по легколетучему компоненту

$$x_i = (Vy_{i-1} + Wx_W)/(V + W) \qquad i < f \qquad \text{(II, 137)}$$

$$x_i = [Vy_{i-1} + (1 - q)\,Fy_p - Dx_D]/[V + (1 - q)\,F - D] \qquad i = f \qquad \text{(II, 138)}$$

$$x_i = (Vy_{i-1} - Dx_D)/(V - D) \qquad i > f \qquad \text{(II, 139)}$$

Уравнение фазового равновесия

$$y_i^* = \alpha x_i/[1 + (\alpha - 1)\,x_i] \qquad \text{(II, 140)}$$

Состав паровой фазы с учетом к. п. д. тарелки

$$y_i = y_{i-1} + (y_i^* - y_{i-1})\,\eta \qquad \text{(II, 141)}$$

Здесь i — номер тарелки; f — номер тарелки питания; F — расход питания, кмоль/ч; D — отбор дистиллята кмоль/ч; W — отбор кубового продукта, кмоль/ч; V — поток пара в колонне, кмоль/ч; x_F, x_D, x_W — содержание легколетучего компонента в питании, дистилляте, кубовом продукте, мол. доли; x_i — содержание легколетучего компонента в жидкой фазе на i-той тарелке, мол. доли; y_i — содержание легколетучего компонента в паровой фазе на i-той тарелке, мол. доли; y_p — содержание легколетучего компонента в паровой фазе питания, мол. доли; y_i^* — равновесное содержание легколетучего компонента в паровой фазе на i-той тарелке; q — доля жидкой фазы в питании; α — относительная летучесть компонентов; η — к. п. д. тарелки. Для расчета q, y_p используется уравнение (II, 146) (см. модель сепаратора).

В зависимости от постановки задачи в качестве независимых переменных может быть принята любая пара из следующих величин: x_W, x_D, D, V. Определение оставшихся неизвестных величин, при заданных значениях независимых переменных сводится к итерационному решению системы уравнений (II, 135)—(II, 141). Значения температур дистиллята и кубового продукта, необходимые для расчета теплового баланса ХТС, определяются из решения уравнения изотерм [44, с. 58]

для кубового продукта

$$\sum_j k_j(T)\, x_{jW} - 1 = 0 \qquad \text{(II, 142)}$$

для дистиллята

$$\sum_j k_j(T)\, x_{jD} - 1 = 0 \qquad \text{(II, 143)}$$

где $k_j(T)$ константа фазового равновесия j-той компоненты, являющаяся функцией температуры; x_{jW}, x_{jD} — содержание j-той компоненты в кубовом продукте и дистилляте, мол. доли.

Колонна азеотропной осушки 5. Предполагается, что расходы и составы исходной и осушенной *н*-пентановой фракции одинаковы. Температура осушенной *н*-пентановой фракции, покидающей куб колонны, рассчитывается по уравнению изотермы (II, 142).

Адсорбер 14. Предполагается, что все параметры (за исключением давления, поступающего на осушку в адсорбер водородсодержащего газа), не претерпевают изменений. Давление на выходе из адсорбера рассчитывается по формулам [44, с. 214]:

$$p_{\text{вых}} = p_{\text{вх}} - \frac{\xi \gamma w^2}{2g} \qquad \xi = \frac{k \lambda H}{d_з} \qquad w = \frac{G}{\gamma 3600 F} \qquad \text{(II, 144)}$$

$$\lambda = \frac{16}{\text{Re}^{0,2}} \quad \text{при} \quad \text{Re} = \frac{w d_з}{\nu} > 40 \qquad \text{(II, 145)}$$

где $p_{\text{вых}}$, $p_{\text{вх}}$ — давление на выходе и входе адсорбера, МПа; ξ — коэффициент сопротивления; $k = 27{,}8$ — коэффициент, зависящий от диаметра зерна; H — высота слоя цеолитов, м; $d_з$ — диаметр зерна, м; w — линейная скорость газа, м/с; γ — удельный вес газа, кгс/м^3; G — расход газа, кг/ч; F — площадь свободного сечения колонны, м^2; g — ускорение свободного падения, м/с^2; ν — кинематическая вязкость газа, м^2/с; λ — коэффициент трения, зависящий от критерия Рейнольдса.

Сепараторы 14, 21, 24. Математическая модель сепаратора предназначена для расчета количеств и составов жидкой и газовой фаз при заданных значениях состава, температуры и давления потока, поступающего на разделение в сепаратор. Для этого необходимо определить корень уравнения [44, с. 71]:

$$f(q) = \sum_i \frac{z_i}{1 - q\left(1 - k_i^{-1}\right)} - 1 = 0 \qquad \text{(II, 146)}$$

где q — искомая доля жидкой фазы; z_i — мольная доля i-той компоненты исходной смеси; $k_i = k_i(t, p)$ — константа фазового рав-

новесия i-той компоненты, являющаяся функцией температуры и давления. Корень q уравнения (II, 146) лежит в пределах $0 \leqslant q \leqslant 1$. Если при $q = 1$ $f(q) < 0$, то весь поток, поступающий на разделение, находится в жидкой фазе. Если $0 < q < 1$, то

$$y_i = z_i\left[1 - q\left(1 - k_i^{-1}\right)\right]^{-1} \qquad x_i = y_i/k_i \qquad L = qF \qquad G = F - L$$

Если $q = 0$, то $y_i = z_i$, $G = F$. Если $q = 1$, то $x_i = z_i$, $L = F$; здесь x_i, y_i — мольные доли i-той компоненты в жидкой и газовой фазах соответственно; F, L, G — количества исходной, жидкой и газовой фаз соответственно.

Компрессор 18. При заданной степени сжатия компрессора, температура на выходе из него вычисляется по формуле [44, с. 48]:

$$T_{\text{вых}} = T_{\text{вх}}\left(\frac{p_{\text{вых}}}{p_{\text{вх}}}\right)^{\frac{k-1}{k\eta_p}} \qquad \frac{1}{k-1} = \sum_i \frac{x_i}{k_i - 1} \tag{II, 147}$$

где $T_{\text{вх}}$, $T_{\text{вых}}$ — температуры на входе и выходе компрессора, К; $p_{\text{вх}}$, $p_{\text{вых}}$ — давление на входе и выходе компрессора, мПа; η_p — политропный к. п. д. компрессора; k — показатель адиабаты; k_i — показатель адиабаты для i-той компоненты смеси; x_i — содержание i-той компоненты в смеси, % (масс.)

Смесители потоков 2, 6, 7, 17, 20. Количество и состав выходного потока рассчитывается по формулам:

$$G_{\text{вых}} = \sum_j G_{\text{вх}\,j} \qquad x_{\text{вых}\,i} = \sum_j \left(G_{\text{вх}\,j} x_{\text{вх}\,j,\,i}/G_{\text{вых}}\right) \tag{II, 148}$$

где $G_{\text{вх}\,j}$ — количество j-го потока, кг/ч; $x_{\text{вх}\,j,\,i}$ — содержание i-той компоненты в j-том потоке, % (масс.). Температура выходного потока рассчитывается итерационно по уравнению

$$t_{\text{вых}} = \frac{\sum_j t_{\text{вх}\,j} G_{\text{вх}\,j} c_{\text{вх}\,j}}{G_{\text{вых}} c_{\text{вых}}} \tag{II, 149}$$

где $t_{\text{вх}\,j}$ — температура j-го входного потока, °С; $c_{\text{вх}\,j}$ — теплоемкость j-го входного потока, кДж/(кг·°С); $c_{\text{вых}}$ — теплоемкость выходного потока, которая является функцией выходной температуры.

Делитель потока 19, 8. Количества, составы и температуры выходных потоков рассчитываются по формулам:

$$G_{\text{вых}\,j} = R_j G_{\text{вх}} \qquad x_{\text{вых}\,j,\,i} = x_{\text{вх}\,i} \qquad t_{\text{вых}\,j} = t_{\text{вх}} \tag{II, 150}$$

где R_j — коэффициент делителя для j-го выходного потока, $\sum_{j=1}^{m} R_j = 1$; $x_{\text{вх}\,i}$ — содержание i-той компоненты во входном потоке, % (масс.); $x_{\text{вых}\,j,\,i}$ — содержание i-той компоненты в j-том выходном потоке, % (масс.); m — число выходных потоков.

Делитель потока 8 рассчитывается после определения $G^{\text{т}}$ для рекуператора 1.

Теплообменники 1, 9, 11, 16, 22, 23, 25. Математическая модель рассматриваемых аппаратов описывается следующей системой уравнений [45, с. 113]:

$$A = G^{т}c^{т}\eta_{п}^{т}/\left(G^{х}c^{х}\eta_{п}^{х}\right) \qquad z = \sqrt{(1+A)^2 - 4pA} \qquad \text{(II, 151)}$$

$$s = \frac{kF}{G^{т}c^{т}\eta_{п}^{т}} \qquad \Phi = \frac{\exp(sz) - 1}{0{,}5\,(z + A + 1) + z} \qquad \text{(II, 152)}$$

$$t_{вых}^{т} = t_{вх}^{т} - \Phi\left(t_{вх}^{т} - t_{вх}^{х}\right) \qquad \text{(II, 153)}$$

$$t_{вых}^{х} = t_{вх}^{х} + \Phi A\left(t_{вх}^{т} - t_{вх}^{х}\right) \qquad \text{(II, 154)}$$

где A, z — функции водяных эквивалентов; s — число единиц переноса тепла; Φ — функция тепловой эффективности; $t_{вых}^{т}$, $t_{вх}^{т}$ — конечная и начальная температуры теплового теплоносителя, °С; $t_{вых}^{х}$, $t_{вх}^{х}$ — конечная и начальная температуры холодного теплоносителя, °С; $G^{т}$, $G^{х}$ — расходы теплого и холодного теплоносителей, кг/ч; $c^{т}$, $c^{х}$ — теплоемкость теплого и холодного теплоносителей, кДж/(кг·°С); $\eta_{п}^{т}$, $\eta_{п}^{х}$ — коэффициенты потерь теплового и холодного теплоносителей; k — коэффициент теплопередачи, кДж/(м²·ч·°С); F — площадь теплопередачи, м²; p — индекс противоточности (для теплообменников *1*, *9*, *11*, *16*, *22*, *25* с противоточной схемой $p=1$, для аппарата *23* с перекрестным током и разделением одной среды на отдельные струи $p = 0{,}57$).

Для аппаратов *9*, *11*, *22*, *23*, *25* при заданных значениях начальных температур и расходов теплоносителей неизвестными являются конечные температуры теплоносителей, которые легко рассчитываются по приведенным формулам. Для обеспечения регламентного режима работы колонны азеотропной осушки *5* и адсорбера *14* в аппаратах *1*, *16* необходимо поддерживать заданную конечную температуру холодного и теплого теплоносителей соответственно. С этой целью для аппарата *1* рассчитывается расход теплого теплоносителя по уравнениям (II, 151), (II, 152), (II, 154), а для аппарата *16* расход холодного по уравнениям (II, 151)—(II, 154). Задача решается итерационно. После расчета $G^{т}$ для аппарата *1* по уравнению (II, 153) рассчитывается $t_{вых}^{т}$. Если для рекуператора *1* рассчитанное значение $G^{т}$ больше потока, поступающего в делитель *8*, то $G^{т}$ принимается равным этому потоку и по уравнениям (II, 151)—(II, 154) вычисляются конечные температуры обоих теплоносителей, соответствующие принятому расходу.

Пароперегревательная печь 12. Математическая модель этой печи описывается уравнением

$$Gc\,(t_{вых} - t_{вх}) = G_{т.г}\,\Delta H\eta \qquad \text{(II, 155)}$$

где G — расход смеси в печь, кг/ч; c — теплоемкость смеси, кДж/(кг·°С); $t_{вых}$, $t_{вх}$ — температура на выходе и входе печи, °С; $G_{т.г}$ — расход топливного газа в печь, кг/ч; ΔH — теплотворная способность топлива, кДж/кг; η — к. п. д. печи.

Постановка задач расчета. При расчете и оптимизации стационарных режимов ХТС важным вопросом является выбор незави-

симых поисковых переменных. При традиционном подходе в качестве независимых поисковых переменных принимаются управляющие переменные аппаратов системы; для ректификационных колонн — это расход пара V и отбор дистиллята D, в теплообменниках — расход хладоагента, в печи — расход топливного газа. Расчет стационарного режима сводится к расчету материального и теплового балансов ХТС при заданных значениях управляющих переменных. Однако в данной постановке возможны случаи, когда на некотором итерационном шаге расчета ХТС, особенно это касается методов, в которых наблюдается колебательная сходимость, отсутствует решение для отдельного аппарата. К примеру, сильное уменьшение потока сырья, поступающего в реакторный узел при заданном значении поисковой переменной (расхода топливного газа в печь), может привести к значительному возрастанию температуры этого потока. Это может явиться причиной отсутствия решения для модели реактора. При фиксированных значениях поисковых переменных (отбор дистиллята и расход пара в колонне), сильное изменение параметров питания, поступающего на разделение, может привести к отсутствию решения для модели ректификации. Корректировка значений расхода топливного газа в печь, отбора дистиллята, расхода пара в колонну с целью получения решения затруднительна и может быть достигнута только методом проб и ошибок. Этих затруднений можно избежать, если в качестве поисковых переменных принять некоторые фазовые переменные. В качестве поисковой переменной модели реакторного узла целесообразно принять температуру на входе в реактор, на которую наложены двухсторонние ограничения, обеспечивающие условия физической реализуемости процесса изомеризации *н*-пентана, в модели ректификации — концентрации легколетучего компонента в дистилляте и кубовом продукте. Тогда при заданных x_F, F, t_F необходимо определить D, V, обеспечивающие заданные составы верхнего x_{D3} и нижнего x_{W3} продуктов. Отбор дистиллята, обеспечивающего заданные составы продуктов разделения, вычисляется из уравнения материального баланса:

$$D = F\,(x_F - x_{W3})/(x_D - x_{W3}) \qquad \text{(II, 156)}$$

Количество пара, необходимое для обеспечения заданных составов продуктов разделения, определяется из потарелочного расчета итерационным решением уравнений (II, 137)—(II, 141), которые обобщенно записываются в виде:

$$x_{D3} - x_D\,(F,\ V,\ x_F,\ t_F) = 0 \qquad \text{(II, 157)}$$

Если на некотором итерационном шаге расчета ХТС окажется, что в уравнении (II, 156) $x_{W3} > x_F$, а следовательно физически реализуемое решение отсутствует, этот недостаток можно легко скорректировать, приняв для данного шага, что $x_{W3} < x_F$.

Замена традиционных поисковых переменных некоторыми фазовыми переменными приводит к следующей постановке задачи рас-

чета стационарных режимов ХТС. При заданных значениях поисковых переменных необходимо рассчитать материальный и тепловой балансы ХТС, а также управляющие воздействия (расход топливного газа в печь, расходы пара и отборы дистиллятов в ректификационных колоннах, расходы теплового теплоносителя в рекуператор *1* и холодного в холодильник *16*), которые обеспечат стационарный режим. Напомним, что для рекуператора *1* в качестве поисковой переменной принята температура холодного теплоносителя, а для холодильника *16* — температура теплого.

Обсуждение результатов расчета. Для расчета стационарного режима схема приведена к условно разомкнутому виду. Размерность вектора разрываемых потоков равна 9. На рис. 8 места разрыва потоков обозначены римскими цифрами. Элементами вектора для потока *I* являются покомпонентные расходы и температуры, для потоков *II*—*IV* — температуры. Были использованы методы простой итерации (PRIT), доминирующего собственного значения (DEM), модифицированный метод Вольфа (WOLF), квазиньютоновский метод (QNM), обобщенный метод доминирующего собственного значения (GDEM) [23].

При расчете ХТС методом PRIT решение было получено примерно за 1000 итераций, что составляло около 30 минут машинного времени ЭВМ ЕС-1033. При столь больших затратах машинного времени на расчет одного стационарного режима ни о какой оптимизации режимов говорить не приходится. Поскольку в моделях ректификации при расчете одной итерации основное время затрачивается на вычисление расхода по уравнению (II, 157), для сокращения времени счета был применен следующий прием. До полного сведения материального и теплового балансов системы в моделях ректификации рассчитывались отборы дистиллята D и кубового продукта W. В точке решения по уравнению (II, 157) вычислялось значение V, соответствующее заданному качеству продуктов разделения. Аналогичным образом, расходы теплого теплоносителя в рекуператор *1* и холодного в холодильник *16*, [рассчитываемые итерационно по уравнениям (II, 151)—(II,154)], обеспечивающие заданные температуры, также рассчитывались только после сведения материального и теплового балансов. Значение неизвестной выходной температуры теплого теплоносителя в рекуператоре *1* до полного расчета схемы не играет роли, так как в уравнении (II, 156) модели ректификации, используемом на каждой итерации, агрегатное состояние питания не учитывается. Описанный подход позволил сократить время расчета схемы более чем на 30 %.

При расчете схемы методами WOLF, QNM, DEM возникали случаи, когда прогнозируемые значения итерируемых переменных принимали отрицательные значения. Чтобы избежать этого, была произведена замена итерируемых переменных. Так, если прогнозируемое значение итерируемой переменной обозначить через z, новая переменная принимает значение $x = |z|$.

Результаты расчета ХТС изомеризации *н*-пентана приведены в таблице 6.

Таблица 6. **Результаты расчета ХТС изомеризации *н*-пентана**

Метод	PRIT	DEM	QNM	WOLF	GDEM
Число итераций	977	59 (4)	7 (17)	23	45 (12)

В строке «число итераций» для методов DEM и GDEM в скобках указано число шагов ускорения, для метода QNM — число вычислений разомкнутой схемы. Начальная невязка (II, 17) составляла $1{,}49 \cdot 10^4$. Решение считалось достигнутым, если величина невязки составляла менее 0,01. Итерационный процесс расчета стационарного режима ХТС методом PRIT был монотонным, вследствие чего работа метода DEM при значении критерия принятия шага ускорения равном 10^{-3} оказалась достаточно эффективной.

Методы QNM и WOLF дали приблизительно одинаковые результаты. На рис. 9 приведены графики зависимости невязок от номера итерации. В случае применения метода GDEM стрелками указаны шаги ускорения. Размерности базисов в случае применения методов WOLF, GDEM были приняты равными 9 и 6 соответственно. При методе WOLF базис набирался чередованием простой итерации и шага по чистому методу Вольфа [1, с. 35]. При завершенном базисе заменялась точка, обладающая максимальной невязкой. Для метода GDEM шаг ускорения принимался, когда значение критерия принятия шага ускорения не превышало 10^{-7}. При методе QNM в качестве начального приближения к матрице Якоби была принята его разностная аппроксимация. Выбор поисковых переменных и связанные с ним подходы в случае расчета стационарного режима

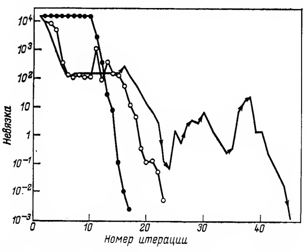


Рис. 9. Зависимость невязки от номера итераций, полученная с помощью методов: ○ — *WOLF*; ● — *QNM*; → — *GDEM*.

ХТС изомеризации *н*-пентана, а также использование методов ускорения сходимости позволили сократить число вычислений схемы примерно в 50 раз, а время решения с 30 до 3 минут.

Для обеспечения заданных производительности установки по изопентану $G_{и\text{-}C_5}$ и мольного соотношения расходов водорода и *н*-пентана на входе в реактор M_3 воспользуемся методом QNM и алгоритмом, учитывающим эти ограничения на одном итерационном уровне расчета схемы. Для этого наряду с итерируемыми переменными (вектором разрываемых потоков x) ищутся такие значения расхода сырья (*н*-пентана) $G_{н\text{-}C_5}$ и водорода G_{H_2} на установку, которые обеспечат выполнение равенств $G_{и\text{-}C_5} = G^3_{и\text{-}C_5}$, $M = M_3$. Решаемая система имеет вид:

$$f_i(x,\ G_{н\text{-}C_5},\ G_{H_2}) = x_i - F_i(x,\ G_{н\text{-}C_5},\ G_{H_2}) = 0,\ i = \overline{1,9} \qquad (II,\ 158)$$

$$f_{10}(x,\ G_{н\text{-}C_5},\ G_{H_2}) = G_{и\text{-}C_5} - G^3_{и\text{-}C_5};\ f_{11}(x,\ G_{н\text{-}C_5},\ G_{H_2}) = M - M_3 = 0 \qquad (II,\ 159)$$

Решение этой системы квазиньютоновскими методами (в частности методом QNM) возможно в случае использования в качестве начальной оценки матрицы Якоби его разностной аппроксимации.

Начальная невязка составляла $1{,}55 \cdot 10^4$. Решение методом QNM было достигнуто всего за 6 итераций. При этом число вычислений разомкнутой схемы было равно 18. Из полученных результатов очевидно, что несмотря на увеличение размерности решаемой системы за счет учета ограничений типа равенств на одном итерационном уровне число вычислений практически не изменилось, так что описанный подход значительно эффективнее многоуровневых алгоритмов типа «цикл в цикле» [3, с. 28].

Расчет больших систем

ХТС средних размеров обычно имеется 20—30 аппаратов, а каждый поток в среднем характеризуется 3—5-ю компонентами. Поэтому система (II, 1), (II, 3) будет иметь размерность 60—150 [46]. Желание более точно моделировать ХТС приводит к тому, что размерность решаемых систем нелинейных уравнений растет. В то же время, число итераций даже для решения линейных систем уравнений растет пропорционально их размерности.

Можно предположить, что и для нелинейных систем число итераций будет расти с увеличением размерности задачи. Поэтому решение системы (II, 1), (II, 3) может потребовать много времени. Особенно это касается случая, когда при оптимизации ХТС приходится многократно рассчитывать ее стационарные режимы для различных значений управляющих переменных. В связи с этим большое значение приобретает разработка эффективных методов решения систем нелинейных уравнений большой размерности.

Рассмотрим подходы к решению задач расчетом ХТС большой размерности. Заметим, что уменьшение размерности не должно быть самоцелью. Главное — это уменьшение продолжительности счета.

В связи с тем, что модели ХТС достаточно сложны, с некоторым приближением можно считать, что время, затраченное на решение, пропорционально числу итераций при решении системы (II, 1), (II, 2).

Последовательный подход. Вначале рассмотрим эту проблему применительно к последовательному подходу. Здесь уменьшение размерности задачи расчета ХТС достигается методами структурного анализа [47]. При этом решаются следующие задачи: 1) в схеме выделяются комплексы — совокупности блоков охваченных обратными связями [3, с. 33]; 2) определение внутри каждого комплекса оптимальной с точки зрения какого-либо критерия совокупности итерируемых переменных (II, 5). Обычно совокупность итерируемых переменных (II, 5) выбирается из условия, чтобы их суммарная размерность была минимальной. Положительные и отрицательные стороны такого выбора переменных (II, 5) обсуждаются в работе [3, с. 85]. Отметим здесь, что применительно к квазиньютоновским методам это более или менее оправдано, поскольку, как мы уже отмечали, можно считать при применении этих методов, что число итераций растет пропорционально размерности системы нелинейных уравнений. Уменьшаются требования и к размеру памяти, поскольку приходится хранить одну или две матрицы размерности $n \times n$. При использовании ориентированного на уравнения подхода так же, как и в предыдущем случае определяются комплексы, а внутри комплексов — оптимальные совокупности разрываемых потоков [48; 17; 18, с. 258].

Параллельный подход. Рассмотрим теперь параллельные методы расчета ХТС применительно к системам (I, 1), (I, 6), или эквивалентной ей системы (II, 4). Непосредственное применение квазиньютоновского метода 2-го рода для решения системы (II, 4) потребует хранения двух $(\overline{N} \times \overline{N})$-матриц, а число итераций, когда все модели линейны, будет равно $\overline{N} = mN$. Поскольку обычно $\overline{N} \gg n$, на первый взгляд может показаться, что переход к параллельному способу только ухудшит результаты. Однако, как показано ниже, использование особенностей структуры ХТС, а, следовательно, особенностей структуры системы (II, 4) может сделать параллельный метод существенно более эффективным.

Особенности структуры ХТС. Для ХТС характерна ситуация, когда каждый блок связан с небольшим числом других блоков. Это значит, что в отдельное уравнение системы (II, 4) будет входить только небольшая часть общего числа переменных $x^{(k)}$, $(k = \overline{1, N})$. Если в i-тое уравнение системы (II, 4) переменная $x^{(k)}$ не входит, то соответствующая частная производная от F_i по этой переменной тождественно равна нулю. Таким образом, матрица Якоби системы (II, 4) будет содержать большое число нулей, и мы имеем дело с системами, обладающими разреженной структурой.

Другая особенность состоит в том, что система (II, 4) имеет блочную структуру. Это особенно заметно, если рассматривать не систему (II, 4), а систему (I, 1), (I, 6), из которой она была получена. На рис. 10 приведен вид матрицы Якоби для систем урав-

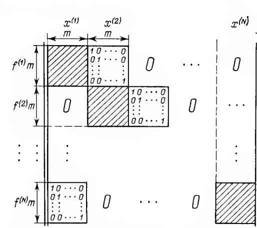


Рис. 10. Вид матрицы Якоби системы уравнений (II,4), описывающей схему, приведенную на рис. 5.

нений (II, 4) применительно к схеме, показанной на рис. 5. Первые m столбцов соответствуют входным переменным 1-го блока, вторые m столбцов — входным переменным второго блока и т. д.; первые m строк — уравнениям 1-го блока, вторые m строк — уравнениям 2-го блока и т. д. Заштрихованные участки соответствуют ненулевым элементам, незаштрихованные части — нулевым элементам.

Следующая особенность состоит в том, что многие элементы матрицы Якоби имеют известные постоянные значения. Так, многие элементы в матрице Якоби, показанной на рис. 10, равны единице.

И, наконец, последняя особенность состоит в том, что матрицы Якоби правых частей уравнений (I, I) для многих блоков легко могут быть вычислены.

При построении квазиньютоновских методов желательно учесть как можно больше свойств самой матрицы Якоби. Общим для этих методов является то, что строится аппроксимация самой матрицы Якоби, а не обратной. Это связано с тем, что сама матрица может иметь большое число нулевых и постоянных элементов, в то время как обратная обычно является заполненной, имеющей мало нулевых и постоянных элементов. Идея построения квазиньютоновских методов, учитывающих разреженную структуру систем нелинейных уравнений, состоит в том, чтобы при построении матриц B_i ($i = 1, 2, \ldots$) сохранена была структура самих матриц Якоби, т. е. если некоторый элемент матрицы Якоби равен нулю, то и соответствующий элемент матрицы B_i должен быть равен нулю. То же требование относится к случаю, когда элемент матрицы Якоби является либо легко вычисляемым, либо постоянным. Все дальнейшее изложение будет вестись применительно к параллельному методу расчета ХТС.

Квазиньютоновские методы 1 рода для решения разреженных систем нелинейных уравнений

Обозначим через M_1 множество пар целых чисел (i, j) таких, что соответствующий элемент $j_{ij}^{(k)}$ матрицы J_k равен некоторой постоянной p_{ij}, не зависящей от точки, а следовательно, и номера итерации (в частности p_{ij} может быть равно нулю), т. е.

$$j_{ij} = p_{ij}, \qquad (i, j) \in M_1 \qquad \text{(II, 160)}$$

Обозначим через M_2 множество пар (i, j) таких, что соответствующие элементы $j_{ij}^{(k)}$ легко вычисляются. Введем множества M, $\overline{M}$:

$$M = M_1 \cup M_2, \qquad \overline{M} = \{(i, j) : (i, j) \notin M_1\}$$

Наша задача состоит в построении такой матрицы B_{k+1}, элементы $b_{ij}^{(k+1)}$, $(k = 0, 1, \ldots)$ которой удовлетворяют соотношениям:

$$b_{ij}^{(k+1)} = \begin{cases} p_{i,j} & (i, j) \in M_1 \\ j_{ij}^{(k+1)} & (i, j) \in M_2 \end{cases} \qquad \text{(II, 161)}$$

Матрицу B_{k+1} будем искать в виде (II, 38). Ясно, что элементы e_{ij} $[(i, j) \in M]$ этой матрицы удовлетворяют соотношениям:

$$e_{ij} = \begin{cases} 0, & (i, j) \in M_1 \qquad \text{(II, 162)} \\ j_{ij}^{(k+1)} - j_{ij}^{(k)} = \Delta j_{ij}^{(k)} & (i, j) \in M_2 \qquad \text{(II, 163)} \end{cases}$$

Для определения матрицы B_{k+1} опять применим принцип наименьшего изменения матрицы B_k. Тогда задача определения матрицы E будет выглядеть следующим образом:

$$\min_{e_{ij}} \frac{1}{2} \sum_{i=1}^{\overline{N}} \sum_{j=1}^{\overline{N}} e_{ij}^2 \qquad \text{(II, 164)}$$

$$\sum_{j=1}^{\overline{N}} e_{ij} s_{j,k} = r_{i,k} \qquad j = \overline{1, \overline{N}} \qquad \text{(II, 165)}$$

$$e_{ij} = 0 \qquad (i, j) \in M_1 \qquad \text{(II, 166)}$$

$$e_{ij} = \Delta j_{ij}^{(k)} \qquad (i, j) \in M_2 \qquad \text{(II, 167)}$$

где $r_{i,k}$ i-тая компонента вектора r_k, определяемого формулой (II, 40). Рассмотрим вначале случай, когда $M_2 = \varnothing$. Преобразуем условия (II, 165). Вектор-строку $s(i)$ с элементами $s(i)_j$ $[\bar{s}(i) = s(i)_1, \ldots, s(i)_{\overline{N}})]$ введем следующим образом:

$$s(i)_j = \begin{cases} s_{j,k} & (i, j) \in \overline{M} \\ 0, & (i, j) \in M \end{cases} \qquad \text{(II, 168)}$$

Подставляя в уравнения (II, 165) значения e_{ij}, удовлетворяющие условию (II, 162), получим

$$\sum_{j \in \overline{M}} e_{ij} s_{j,k} = r_{i,k} \qquad \text{(II, 169)}$$

Если использовать вектор $s(i)$, эти соотношения могут быть переписаны в виде:

$$\sum_{j=1}^{\overline{N}} e_{ij} s(i)_j = r_{i,k} \qquad i = \overline{1, \overline{N}} \qquad \text{(II, 170)}$$

Как видно из дальнейшего, запись условий (II, 165) в виде (II, 170) обеспечивает автоматическое выполнение условий (II, 166), поэтому соотношения (II, 170) эквивалентны соотношениям (II, 165), (II, 166). Заменяя в задаче (II, 164)—(II, 166) условия (II, 165), (II, 166)

соотношениями (II, 170) и учитывая, что $M_2 = \varnothing$, сводим эту задачу к следующей:

$$\min_{e_{ij}} \sum_{i=1}^{\overline{N}} \sum_{j=1}^{\overline{N}} e_{ij}^2 \quad \sum_{j=1}^{\overline{N}} e_{ij} s(i)_j = r_{i,k} \tag{II, 171}$$

Соответствующая функция Лагранжа имеет вид

$$L = \frac{1}{2} \sum_{i=1}^{\overline{N}} \sum_{j=1}^{\overline{N}} e_{ij}^2 + \sum_{i=1}^{\overline{N}} \lambda_i \left[\sum_{j=1}^{\overline{N}} e_{ij} s(i)_j - r_{i,k} \right] \tag{II, 172}$$

Приравнивая производные от L по элементам e_{ij} нулю, получим

$$e_{ij} = -\lambda_i s(i)_j \tag{II, 173}$$

Поскольку $s(i)_j = 0$ для $(i, j) \in M_1$, то $e_{ij} = 0$ для $(i, j) \in M_1$.

Подставим e_{ij} из формулы (II, 173) в выражение (II, 170), получим

$$\lambda_i = -\frac{r_{i,k}}{s_i(i)^{\mathrm{T}} s(i)}$$

Отсюда

$$e_{ij} = r_{i,k} \frac{s(i)_j}{s(i)^{\mathrm{T}} s(i)} \tag{II, 174}$$

Обозначим через e_i строчку матрицы E, тогда

$$e_i = r_{i,k} \frac{s(i)}{s(i)^{\mathrm{T}} s(i)} \tag{II, 175}$$

Обозначим через $b_i^{(k)}$ i-тую строчку матрицы B_k; тогда из уравнения (II, 38), используя формулы (II, 40), (II, 10), (II, 22), получим:

$$b_i^{(k+1)} = b_i^{(k)} + \frac{[f_{i,k+1} - (1 - \alpha_k) f_{i,k}] s(i)}{s(i)^{\mathrm{T}} s(i)} \tag{II, 176}$$

Эта формула выведена Шубертом [49], но другим путем. Ясно, что если $(i, j) \in M_1$, то $e_{ij} = 0$ в соответствии с (II, 168). Поэтому $b_{ij}^{(k+1)} = b_{ij}^{(k)}$ для $(i, j) \in M_1$. Аналогично можно показать, что $b_{ij}^{(k)} = b_{ij}^{(k-1)}$ для $(i, j) \in M_1$, и т. д. Таким образом, если в начальной матрице B_0 элементы $b_{ij}^{(0)} = p_{ij}$ для $(i, j) \in M_1$, то и для всех последующих матриц соответственные элементы будут равны величинам p_{ij}.

Рассмотрим теперь случай, когда множество M_2 не пусто. В этом случае соотношение (II, 169) будет иметь вид:

$$\sum_{(i,j) \in \overline{M}} e_{ij} s_{j,k} + \sum_{(i,j) \in M_2} e_{ij} s_{j,k} = r_{i,k} \tag{II, 177}$$

Ясно, что вторая сумма в левой части равенства (II, 177) является известной величиной [см. выражение (II, 163)]. Запишем равенство (II, 177) в виде:

$$\sum_{(i,j) \in \overline{M}} e_{ij} s_{j,k} = \bar{r}_{i,k} \tag{II, 178}$$

где $\tilde{r}_{i,k} = r_{i,k} - \sum_{(i,j) \in M_2} \Delta j_{ij}^{(k)} s_{j,k}$ при $i = \overline{1, N}$ — известные величины. Введем новые переменные $\tilde{e}_{ij}$

$$\tilde{e}_{ij} = \begin{cases} 0 & (i, j) \in M \\ e_{ij} & (i, j) \in \overline{M} \end{cases} \qquad (II, 179)$$

Соотношения (II, 178) с учетом выражений (II, 168), (II, 179) могут быть переписаны в виде:

$$\sum_{j=1}^{\overline{N}} \tilde{e}_{ij} s(i)_j = \tilde{r}_{i,k} \qquad (II, 180)$$

По построению соотношение (II, 180) эквивалентно соотношениям (II, 165)—(II, 167). Заменяя в задаче переменные e_{ij} на переменные $\tilde{e}_{ij}$, а соотношения (II, 165)—(II, 167) — соотношениями (II, 180), сведем задачу (II, 164)—(II, 167) к виду:

$$\min_{\tilde{e}_{ij}} \sum_{i=1}^{\overline{N}} \sum_{j=1}^{\overline{N}} \tilde{e}_{ij}^2 \qquad (II, 181)$$

$$\sum_{j=1}^{\overline{N}} \tilde{e}_{ij} \bar{s}(i)_j = \tilde{r}_{i,k}$$

Задача (II, 181) совершенно идентична задаче (II, 171). Поэтому по аналогии с предыдущим случаем [см. выражение (II, 175)] получим:

$$\tilde{e}_{ij} = \frac{\tilde{r}_{i,k} \bar{s}(i)_j}{\bar{s}(i)^{\mathrm{T}} \bar{s}(i)}$$

Отсюда, используя переменные (II, 179), (II, 162), (II, 163), подсчитываем элементы матрицы B_{k+1}:

$$b_{ij}^{(k+1)} = \begin{cases} p_{ij} & (i, j) \in M_1 \\ j_{ij}^{(k+1)} & (i, j) \in M_2 \\ b_{ij}^{(k)} + e_{ij} & (i, j) \in \overline{M} \end{cases} \qquad (II, 182)$$

Полученные формулы (II, 176), (II, 182) являются обобщением формулы Шуберта на случай наличия в матрице B_k постоянных и легко вычисляемых элементов.

Отметим одну особенность. В методе Бройдена рекуррентное соотношение (II, 47) записывается для всей матрицы сразу. В данном же случае рекуррентное соотношение записывается отдельно для каждой строчки матрицы B_k. Это связано с тем, что вектор $s(i)$ [см. соотношение (II, 168)] зависит от номера строки i. После того, как матрица B_k будет определена, необходимо решить систему линейных уравнений (II, 22) для определения направления движения p_k. Система (II, 22) будет системой с разреженной структурой и для ее решения должны быть использованы специальные методы [36].

Один из подходов к решению разреженной системы (II, 22) состоит в следующем. Для системы (II, 22) проводится структурный

анализ, в результате чего находится совокупность итерируемых переменных (II, 5), обладающих минимальной суммарной размерностью. Далее, относительно итерируемых переменных решается система (II, 6), которая в данном случае является линейной [см. выражение (II, 20)]. Вид матрицы A и вектора b неизвестен, и только по заданному $\bar{x}$ можно найти $f(\bar{x}) = A\bar{x} + b$. Для решения системы (II, 20) могут быть использованы формулы (II, 101), (II, 102).

Для ХТС с блоками, описываемыми нелинейными моделями, интересно выяснить, что лучше для разреженной системы нелинейных уравнений (I, 1), (I, 6): провести структурный анализ и свести дело к решению системы нелинейных уравнений (II, 8) и далее применять обычные квазиньютоновские методы, или же подойти к ней, как к системе с разреженной структурой и применять специальные квазиньютоновские методы, а структурный анализ применять для «сжатия» линейной системы (II, 22).

Квазиньютоновский метод с памятью решения разреженных систем нелинейных уравнений

Квазиньютоновский метод с памятью для решения неразреженных нелинейных систем был рассмотрен ранее. При этом элементы матрицы E находились как решение задачи (II,82), (II,83). В данном случае элементы матрицы E будут находиться так же, но при наличии дополнительных условий (II,162). По аналогии с соотношениями (II,168) введем вектор-строку $\bar{s}(i)_m$ с элементами $s(i)_{j,\,m}$, причем сделаем это следующим образом:

$$s(i)_{j,\,m} = \begin{cases} s_{j,\,m} & (i,\ j) \in \overline{M} \\ 0 & (i,\ j) \in M_1 \end{cases} \qquad \text{(II, 183)}$$

где m — номер итерации. По аналогии с соотношением (II,170) соотношения (II,83) могут быть записаны в виде:

$$\sum_{j=1}^{\overline{N}} e_{ij} s(i)_{j,\,m} = r_{i,\,m} \quad m = k,\ k-1,\ \ldots,\ k-q+1$$

или в векторной форме:

$$s(i)_m e_i^{\text{T}} = r_{i,\,m} \qquad m = k,\ k-1,\ \ldots,\ k-q+1 \qquad i = \overline{1,\ N} \qquad \text{(II, 184)}$$

Функция Лагранжа будет иметь вид

$$L = \frac{1}{2} \sum_{i=1}^{\overline{N}} e_i^{\text{T}} e_i + \sum_{p=1}^{q} \sum_{i=1}^{\overline{N}} \lambda_i^{(p)} s(i)_{k-p+1} \cdot e_i^{\text{T}}$$

Приравнивая производные L по e_{ij} нулю, получим

$$e_i = -\sum_{p=1}^{q} \lambda_i^{(p)} s(i)_{k-p+1}$$

Подставляя это выражение для e_i в соотношение (II,184), получим систему $\overline{N}q$ линейных уравнений для определения $\overline{N}q$ величин $\lambda_i^{(p)}$ $(p = 1,\ q;\ i = \overline{1,\ \overline{N}})$.

$$-\sum_{p=1}^{q} \lambda_i^{(p)} s(i)_m s(i)_{k-p+1}^{\text{T}} = r_{i,\,m} \qquad i = \overline{1,\ \overline{N}},\ m = k,\ \ldots,\ k-q+1$$

(II, 185)

Конечно, система (II,185) получилась сложнее системы (II,85). Однако и здесь благодаря специальной структуре можно существенно упростить процедуру решения. Действительно, система (II,185) может быть записана в следующем матричном виде:

$$\begin{aligned} S_{11}\lambda^{(1)} + S_{12}\lambda^{(2)} + \cdots + S_{1q}\lambda^{(q)} &= -r_k \\ \vdots \qquad\qquad & \quad \vdots \\ S_{q1}\lambda^{(1)} + S_{q2}\lambda^{(2)} + \cdots + S_{qq}\lambda^{(q)} &= -r_{k-q+1} \end{aligned} \qquad (II, 186)$$

где $\lambda^{(j)}$ — вектор множителей Лагранжа, соответствующих j-му условию (II,83), а S_{pl} — диагональная $(\overline{N} \times \overline{N})$-матрица. Диагональный элемент s_{ii}^{pl} этой матрицы, стоящий на i-том месте, имеет вид: $s\,(i)_{k-p+1} s\,(i)_{k-l+1}^{\mathrm{T}}$. Матрица системы (II,186) имеет клеточную структуру, каждый элемент ее S_{ij} сам является матрицей.

Предположим, что все матрицы S_{ij} не вырождены; тогда для решения системы (II,186) может быть применен матричный вариант метода исключения Гаусса. При этом, так же, как и в обычном методе исключения, матрица системы (II,186) сначала приводится к треугольной матрице (правда, в данном случае клеточной), после чего начинается процесс определения векторов $\lambda^{(j)}$ $(j = \overline{1, q})$.

Квазиньютоновский метод с блочной аппроксимацией

Данный метод будет основан на использовании блочной структуры системы (I, 1), (I, 6). Откажемся от формирования приближения для всей матрицы Якоби системы (I, 1), (I, 6) сразу и будем строить аппроксимации отдельно для матриц Якоби правых частей каждого из соотношений (I, 1) [50, 26], используя информацию о входных и выходных переменных данного блока, которую получим во время проведения итерационной процедуры решения системы (I, 1), (I, 6). Интересно отметить, что в предыдущем случае также пришлось отказаться от построения аппроксимации для всей матрицы Якоби сразу и перейти к построчной аппроксимации.

Вначале рассмотрим случай, когда размерности всех входных и выходных потоков блоков равны $n_k = m_k = m$. Обозначим через $x_i^{(k)}$, $z_i^{(k)}$ значения векторов $x^{(k)}$ на i-той итерации. (Надеемся, что не возникнет путаницы, связанной с тем, что в остальных главах через $x_i^{(k)}$, $z_i^{(k)}$ обозначались соответственно i-тая входная и выходная переменные k-го блока). Введем обозначения

$$y_i^{(k)} = z_{i+1}^{(k)} - z_i^{(k)} \qquad s_i^{(k)} = x_{i+1}^{(k)} - x_i^{(k)} \qquad (II, 187)$$

Будем считать, что итерируемые переменные являются переменными $x^{(k)}$, $(k = \overline{1, \overline{N}})$. По аналогии с матрицами $Y_i S_i$ [см. выражение (II, 27)] введем матрицы

$$Y_i^{(k)} = \left(y_0^{(k)}, \ldots, y_{i-1}^{(k)}\right) \qquad S_i^{(k)} = \left(s_0^{(k)}, \ldots, s_{i-1}^{(k)}\right) \qquad (II, 188)$$

Обозначим через $B_i^{(k)}$ некоторую аппроксимацию матрицы Якоби правых частей соотношения (I, 1) на i-той итерации. Тогда, при тех же предположениях, при которых было получено матричное уравнение (II, 29) и используя тот же самый вывод, можно получить матричное уравнение для определения $B_i^{(k)}$

$$B_i^{(k)} S_i^{(k)} = Y_i^{(k)} \qquad (II, 189)$$

Для определения $B_i^{(k)}$ из уравнения (II, 189) мы можем воспользоваться формулами (II, 103), (II, 104), либо любым другим аналогом формул (II, 90), (II, 91), выписанным для определения матрицы B_{i+1}. После того, как на i-том шаге будут определены все $B_i^{(k)}$, необходимо найти приращения $\Delta x_i^{(k)}$, $\Delta z_i^{(k)}$, с помощью которых будут определены следующие приближения для $x_i^{(k)}$, $z_i^{(k)}$. Это аналог операции определения Δx_j с помощью уравнения (II, 11) в методе Ньютона и операции определения p_j из уравнения (II, 22) в обычном квазиньютоновском методе.

Из уравнений (I, 1), (I, 6) с точностью до бесконечно малых второго порядка имеем:

$$z_i^{(k)} + \Delta z_i^{(k)} = f^{(k)}\left(x_i^{(k)}\right) + J_i^{(k)} \Delta x_i^{(k)} \qquad \text{(II, 190)}$$

$$\Delta x_i^{(k)} = \sum_{j=1}^{N} \alpha_{kj} \Delta z_i^{(j)} \qquad \text{(II, 191)}$$

В уравнении (II, 190) заменим матрицу Якоби $J_i^{(k)}$ ее аппроксимацией $B_i^{(k)}$:

$$\Delta z_i^{(k)} = B_i^{(k)} \Delta x_i^{(k)} + f^{(k)}\left(x_i^{(k)}\right) - z_i^{(k)} \qquad \text{(II, 192)}$$

Решив систему линейных уравнений (II, 192), (II, 191), найдем выражения $\Delta x_i^{(k)}$, $\Delta z_i^{(k)}$, после чего новые приближения для $x_i^{(k)}$ подсчитаем по формулам:

$$x_{i+1}^{(k)} = x_i^{(k)} + \alpha \Delta x_i^{(k)} \qquad z_{i+1}^{(k)} = z_i^{(k)} + \alpha \Delta z_i^{(k)} \qquad \text{(II, 193)}$$

Система линейных уравнений (II, 191), (II, 192), как правило, имеет разреженную матрицу коэффициентов, поэтому для ее решения могут быть использованы также специальные методы [36].

К решению этой системы можно подойти и по-другому. Действительно, системе уравнений (II, 191), (II, 192) можно дать следующую схемную интерпретацию. Ее можно рассматривать как математическую модель системы, структура которой совпадает со структурой исходной ХТС, а модели блоков [см. соотношения (II, 192)] — линейны. Отсюда, для решения системы (II, 191), (II, 192) удобно использовать последовательный способ расчета ХТС.

Данный метод обладает следующими свойствами.

1. Автоматически учитывает все нулевые элементы матрицы Якоби системы (I, 1), (I, 6), которые находятся вне матриц Якоби $J^{(k)}$, ($k = 1, \overline{N}$) отдельных блоков.

2. Легко позволяет учесть случай, когда матрица Якоби для какого-либо блока либо состоит из постоянных элементов, либо легко рассчитывается. Действительно, пусть для блока k легко рассчитывается его матрица Якоби $J^{(k)}$, тогда в уравнении (II, 192) k-го блока мы будем использовать не $B_i^{(k)}$, а саму матрицу $J^{(k)}$.

3. Обладает m-шаговым свойством линейного окончания (m — размерность одного потока). Действительно, рассмотрим следующий гипотетический случай. Пусть все модели блоков ХТС линейны и имеют вид (II, 3), но виды $A^{(k)}$ и $b^{(k)}$ неизвестны, можно только вы-

числить $z^{(k)}$ по заданному значению $x^{(k)}$. Задача состоит в восстановлении вида $A^{(k)}$ и $b^{(k)}$ на основании информации о $z^{(k)}$ и $x^{(k)}$, которую получим во время итерационного расчета ХТС. Как видно, она очень близка к задаче регрессионного анализа, когда по известным входным и выходным переменным некоторого объекта требуется построить его линейную модель. Так же, как и было доказано равенство (II, 105), в этом случае легко показать, что если векторы $s_i^{(k)}$ $(i = \overline{0, m-1})$ линейно независимы, то на m-том шаге будет выполняться равенство

$$B_m^{(k)} = A^{(k)} \qquad (II, 194)$$

где значения $B_i^{(k)}$ $(i = \overline{1, m})$ подсчитываются по формулам (II, 103), (II, 104). Таким образом, на m-том шаге будут определены все $A^{(k)}$. В этой же точке можно подсчитать вектор $b^{(k)}$ из соотношения (II, 3), поскольку $z^{(k)}$ и $x^{(k)}$ в этой точке известны, а матрица $A^{(k)}$ уже определена. Поскольку после этого в каждом блоке будут определены $A^{(k)}$ и $b^{(k)}$, то решением системы (II, 3), (I, 6) найдем стационарный режим ХТС.

Рассмотрим теперь случай, когда размерности m_k входных и выходных потоков различных блоков не равны. Обозначим через $\bar{k}$, $\bar{\bar{k}}$ номера блоков, для которых m_k принимает соответственно наибольшее и наименьшее значения. Для нахождения матриц $B^{(k)}$, аппроксимирующих матрицы Якоби $J^{(k)}$, используем описанную выше процедуру, а для определения $B^{(k)}$ из уравнения (II, 189) будем использовать формулы (II, 103), (II, 104). Тогда при $i \leqslant m_{\bar{\bar{k}}}$ процесс решения не будет ничем отличаться от описанного выше. При $i = m_{\bar{\bar{k}}}$ будет сформирована матрица Якоби $\bar{\bar{k}}$-го блока. Однако поскольку матрицы Якоби всех остальных блоков еще не будут сформированы, процесс решения должен быть продолжен. Мы видели (см. с. 000), что применение формул (II, 103), (II, 104) при $i > m_{\bar{\bar{k}}}$ будет давать равенство $B_i = J^{(k)}$, $i \geqslant m_{\bar{\bar{k}}}$. Аналогичная ситуация характерна для всех остальных блоков $k \neq \bar{\bar{k}}$. Отсюда при $i = m_{\bar{k}}$ получим точные значения матриц Якоби для всех блоков, а, следовательно, на $(m_{\bar{k}} + 1)$-вом шаге будет получено решение системы (II, 3), (I, 6). Применение метода может натолкнуться на затруднения, если в каком-либо блоке (например, в k-том) последовательные векторы $s_i^{(k)}$ $(i = 0, 1, \ldots, m_k)$ окажутся линейно-зависимыми или близкими к линейно-зависимым. Действительно, матрица $B_i^{(k)}$ определяется из матричного уравнения (II, 189) и, если векторы $s_i^{(k)}$ близки к линейно-зависимым, то матрица $S_i^{(k)}$ будет близка к вырожденной, и определение $B_i^{(k)}$ из уравнения (II, 189) затруднено.

Рассмотрим теперь случай, когда блок может иметь более чем один вход или выход. Этот случай легко сводится к предыдущему. Действительно, рассмотрим, например, случай, когда k-тый блок имеет один входной и два выходных потока, характеризуемые

векторами $z^{(k1)}$, $z^{(k2)}$. Математическая модель блока имеет вид

$$z^{(k1)} = f^{(k1)}(x^{(k)}) \qquad z^{(k2)} = f^{(k2)}(x^{(k)}) \qquad (II, 195)$$

Тогда формально можем считать, что имеются два блока, входные потоки которых характеризуются одним и тем же вектором $x^{(k)}$, а выходные — векторами $z^{(k1)}$, $z^{(k2)}$ соответственно. После этого мы можем применять описанный выше подход, аппроксимируя отдельно матрицы Якоби правых частей равенств (II, 195).

Используем теперь ту же самую гипотетическую схему, что и при рассмотрении свойства 3, для сравнения последовательного подхода с параллельным, при котором используется квазиньютоновский метод с блочной аппроксимацией. В дальнейшем будем называть этот подход параллельным методом. При использовании последовательного метода в сочетании с любым квазиньютоновским методом 2-го рода потребуется n шагов (здесь n — суммарная размерность разрываемых потоков) для определения решения системы (II, 3), (I, 6); при этом потребуется $2n^2$ ячеек памяти для хранения матриц H_i и K_i. При параллельном методе, как мы видели, для определения решения системы (II, 3), (I, 6) потребуется m шагов (m — размерность одного потока). Это очень интересный факт. В данном случае число итераций определяется не общей размерностью системы, которая может быть очень большой (в данном случае она равна $2Nm$), а максимальной размерностью потока (блока). Причем при усложнении структуры ХТС (увеличение числа обратных связей) величина n может существенно возрасти, что в свою очередь приведет к увеличению числа итераций при использовании последовательного метода. В то же время при параллельном подходе число итераций будет определяться только размерностью m одного потока, независимо от сложности структуры ХТС. Конечно, эти выводы верны только для линейных систем, однако подобное свойство рассмотренных методов может проявиться и при решении систем, близких к линейным. Параллельный метод потребует $2Nm^2$ ячеек памяти, поскольку в каждом блоке для определения $B_i^{(k)}$ необходимо использовать две матрицы см. выражения (II, 103), (II, 104). Отсюда ясно, что при $m < n$ и применении параллельного метода число итераций будет меньше. При этом параллельный метод будет требовать меньшего объема памяти, если $m\sqrt{2N} < n$.

Отметим теперь сравнительные достоинства и недостатки; по сравнению с методом Шуберта данный метод обладает следующими преимуществами. Во-первых, он обладает свойством линейного окончания (для метода Шуберта это свойство не доказано). Во-вторых, он может быть применен, когда система (I, 1), (I, 6) не является разреженной, т. е. в случае, когда каждый блок схемы связан с большим числом остальных блоков ХТС. Недостаток же по сравнению с методом Шуберта состоит в том, что, как мы указывали, применение этого метода может столкнуться с трудностями, если в отдельных блоках m последовательных векторов $s_i^{(k)}$ станут близкими к линейно зависимым.

Применение методов решения систем нелинейных уравнений

Обычно необходимо рассчитать стационарный режим при различных значениях управляющих переменных u. Различают два режима расчета системы (II, 6) при изменении переменных u. В первом случае расчет системы (II, 6) проводится для небольшого числа значительно отличающихся одно от другого значений управлений u. Во втором случае проводится многократное решение системы (II, 6) для многих значений вектора u, мало отличающихся одно от другого. Типичный пример такого случая — это решение задачи оптимизации ХТС, когда переменные u меняются в соответствии с некоторой стратегией поиска, и при каждом значении u приходится решать уравнения (II, 6), описывающие стационарный режим схемы. Ко второму случаю сведется также решение систем нелинейных уравнений методом продолжения по параметру, а также решение систем нелинейных уравнений на каждом шаге интегрирования при решении систем обыкновенных дифференциальных уравнений каким-либо неявным методом. Рассмотрим отдельно эти случаи, поскольку учет их специфики может существенно повысить эффективность процедуры расчета системы (II, 6).

Первый случай. Рассмотрим первый режим решения системы (II, 6). В этом случае, как правило, имеется плохое начальное приближение к решению системы (II, 6). В то же время для хорошей работы квазиньютоновских методов необходимо выполнение следующих условий:

1. Матрица H_0 (или B_0) должна быть близка к обратной матрице Якоби (прямой матрице Якоби).

2. Система (II, 6) должна быть близка к линейной; это условие будет выполняться, если начальное приближение находится достаточно близко от решения системы (II, 6). Действительно, при этих условиях шаг в соответствии с (II, 14), (II, 23) будет почти ньютоновским, примененным к системе, близкой к линейной, а, как мы видим, метод Ньютона дает решение системы линейных уравнений за один шаг. При невыполнении этих условий трудно ожидать хорошей сходимости метода. А поскольку при плохом начальном приближении второе условие часто не выполняется, то и метод в этих случаях сходится не очень быстро. И, действительно, типичная картина зависимости нормы правых частей системы от номера итерации проиллюстрирована на рис. 9. Вначале достаточно долго наблюдается очень медленная сходимость, и только в конце итерационного процесса норма начинает очень быстро уменьшаться, т. е. сверхлинейная сходимость появляется только в конце итерационного процесса, когда выполняются оба условия, матрица H_i становится близкой обратной матрице Якоби, а система (II, 6) вследствие близости итерационной точки к точке решения становится близкой к линейной.

Итак, квазиньютоновские методы, как правило, имеют невысокую скорость сходимости, если в качестве начального приближе-

ния используется плохая аппроксимация матрицы Якоби. Поэтому чаще всего в начальной точке определяется разностная аппроксимация матрицы Якоби, для получения которой требуется $n + 1$ вычисление левых частей системы (II, 6). В связи с этим, если методы простой итерации DEM, GDEM сходятся меньше, чем за n итераций, переходить к квазиньютоновским методам невыгодно. С другой стороны, интересно рассмотреть возможность и целесообразность совместной работы различных методов. При этом, конечно, соединение методов не должно быть чисто механическим, но при их соединении информация, накопленная в результате работы одного метода, должна использоваться в другом методе.

В виде примера рассмотрим последовательное применение метода простой итерации и квазиньютоновского метода. Поскольку вначале квазиньютоновский метод часто дает плохую сходимость, в случае, когда метод простой итерации обеспечивает сходимость, может оказаться выгодным вначале на первых n шагах, использовать метод простой итерации, но на каждом шаге векторы x_i и f_i использовать для преобразования матрицы H_i (или B_i) в соответствии с той или иной формулой квазиньютоновского метода, например, по формулам (II, 107), (II, 108). При $i = n$ надо переходить на квазиньютоновский метод, причем в качестве начального приближения к матрице H_0 (или B_0) использовать полученную к этому шагу матрицу H_i. Аналогично может оказаться выгодным применять метод DEM вначале, а затем переходить на квазиньютоновский метод.

Метод Вольфа часто обеспечивает быструю скорость сходимости вначале и медленную в конце. В этом случае также может оказаться полезным сначала применять метод Вольфа, а потом переходить на квазиньютоновский метод. Причем поскольку на i-том ($i > n$) шаге в методе Вольфа [3, с. 35] имеются n предыдущих значений f_j, x_j, ($j = i - 1,\ i - 2, \ldots, i - n + 1$), эти значения могут быть использованы в системе уравнений (II, 32) для определения начального приближения H_0, с которого начнет работать квазиньютоновский метод. В этом случае при переходе от метода Вольфа к квазиньютоновскому методу не понадобится дополнительных вычислений для построения начального приближения H_0.

Второй случай. Рассмотрим теперь второй режим. Пусть приходится находить решение системы (II, 6) для последовательных векторов u_i ($i = 1, 2, \ldots$). Будем исходить из предположения о том, что векторы u_i изменяются на небольшую величину, т. е. выполняется соотношение

$$\| u_{i+1} - u_i \| \leqslant \varepsilon \qquad i = 1, 2, \ldots \tag{II, 196}$$

где ε — малая величина. При определенных условиях (будем считать, что эти условия выполняются) система (II, 6) определяет переменные x как неявные непрерывные функции переменных u. Обозначим через $x_i^* = x_i^*(u_i)$ решение системы (II, 6) при $u = u_i$. Пусть с помощью одного из квазиньютоновских методов найдено решение системы (II, 6) при $u = u_i$. Будем считать, что аппроксимировалась сама матрица Якоби; обозначим через B_i^* предельное

значение матрицы J_i, аппроксимирующей матрицу Якоби. Можно предположить, что матрица B_i^* достаточно близка к матрице J_i. Пусть теперь требуется найти решение системы (II, 6) при $u = u_{i+1}$

$$f(x, u_{i+1}) = 0 \qquad (II, 197)$$

где u_{i+1} удовлетворяет соотношению (II,196). Из соотношения (II,196) и непрерывности функций $x^*(u)$ вытекает, что x_{i+1}^* будет достаточно близко к x_i^* т. е.

$$\| x_{i+1}^* - x_i^* \| < \delta \qquad (II, 198)$$

где δ — малая величина. Отсюда, при решении системы (II, 197) в качестве начальных значений для переменных x целесообразно применять значение x_i^*, полученное на предыдущей итерации

$$x_0 = x_i^* \qquad (II, 199)$$

Это обеспечит хорошее начальное приближение для переменных x. Отсюда мы можем надеяться на быструю сходимость квазиньютоновского метода.

Остановимся теперь на выборе начального приближения для матрицы B при решении системы (II, 197). При выполнении условий (II, 196), (II, 199) и при условии, что кривизна функции f невелика, можно предположить, что матрица B_i^*, полученная на предыдущем шаге, будет хорошим приближением к J_{i+1}, поэтому в качестве B_0 при решении системы (II, 197) может оказаться целесообразным взять матрицу B_i^*. По сравнению с разностной аппроксимацией матрицы Якоби в начальной точке этот прием избавляет нас от дополнительных n расчетов левых частей системы (II, 197) в начальной точке. Хорошее начальное приближение может позволить отказаться от требования, чтобы на каждом направлении при решении системы (II, 197) происходило уменьшение нормы левых частей системы (II, 197), т. е. отказаться от выбора α в выражении (II, 14) из условия (II, 18) или (II, 19), что было вызвано желанием обеспечить стабильность поиска даже при плохом начальном приближении. В данном случае α будет полагаться равным единице как и в ньютоновском методе.

В заключение следует отметить, что квазиньютоновский метод позволяет лучше использовать информацию, полученную на предыдущем шаге, чем методы DEM, GDEM и Вольфа. Действительно, в последних трех методах могут использовать только значения переменных x, полученные на предыдущем шаге [см. выражения (II, 199)], в то время как квазиньютоновский метод использует информацию и о матрице Якоби, полученную на предыдущем шаге. Поэтому в данном случае он может оказаться предпочтительнее упомянутых методов.

Рассмотрим теперь применение метода Ньютона. В этом случае можно улучшить начальное приближение (II, 199), взяв его в виде:

$$x_0 = x_i^* + \frac{\partial x^*}{\partial u}(u_{i+1} - u_i) \qquad (II, 200)$$

где производные $\partial x^*/\partial u$ определяются из (II, 6) по правилу дифференцирования неявных функций.

*Расчет процесса синтеза аммиака**

Синтез аммиака из азотоводородной смеси

$$N_2 + 3H_2 \rightleftarrows 2NH_3 + Q$$

является основным источником связанного азота для химической промышленности и производства минеральных удобрений.

В настоящее время синтез аммиака проводят в крупных агрегатах мощностью 1000—3000 т NH_3 в сутки. Агрегат состоит из отделения подготовки синтез-газа (азотоводородная смесь) при давлении до $30 \cdot 10^6$ Па и собственно отделения синтеза аммиака, проводимого под давлением $(200—300) \cdot 10^6$ Па.

Рассмотрим технологическую схему отделения синтеза аммиака (рис. 11). Циркуляционный газ Γ_1, содержащий a_1 долей аммиака и u_1 инертов (аргон и метан) поступает в колонну синтеза *1*. На выходе из нее газ Γ_2 содержит a_2 долей аммиака. Производительность колонны синтеза

$$\Pi_{кс} = \Gamma_2 a_2 - \Gamma_1 a_1$$

определяется объемом катализатора $O_к$ в колонне.

Образующийся аммиак отделяется в конденсаторах *2* и *6* при температурах $t_{к1}$ и $t_{к2}$ соответственно и выводится из технологического цикла. Инерты частично растворяются в жидком аммиаке, частично выводятся из технологического цикла в составе продувочных газов $\Gamma_{пр}$, доля отбора которых

$$\alpha = \Gamma_{пр}/\Gamma_{цг} \text{ (}\Gamma_{цг} \text{ — поток циркулирующего газа).}$$

является важным режимным параметром технологической схемы. Свежая азотоводородная смесь $\Gamma_{св}$ подается в испаритель *7* вторичного конденсатора *6*. Жидкий аммиак $Ж_1$ и $Ж_2$ из аппаратов *2* и *6* и амиак $Ж_3$ из конденсатора продувочных газов *10* поступает в сборник (танк) *13*. При этом давление сбрасывается до $20 \cdot 10^6$ Па и десорбированные танковые газы, содержащие большой процент аммиака, поступают в конденсатор танковых газов *14*. Конденсат $Ж_1$ возвращается в сборник *13*. Производительность схемы в

Рис. 11. Схема отделения синтеза аммиака:

1 — колонна синтеза; *2* — воздушный конденсатор; *3* — сепаратор; *4* — аммиачно-холодильная установка; *5* — узел отбора продувочных газов; *6* — конденсационная колонна; *7* — испаритель; *8* — аммиачно-холодильная установка; *9* — циркуляционный компрессор; *10* — конденсатор продувочных газов; *11* — испаритель аммиака; *12* — аммиачно-холодильная установка вымораживания танковых и продувочных газов; *13* — сборник жидкого аммиака; *14* — конденсатор.

* Раздел написан совместно с Д. Н. Мотылём.

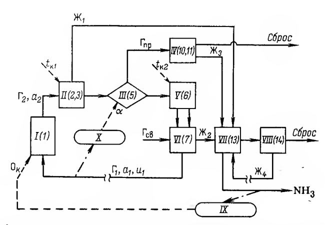


Рис. 12. Информационная схема отделения синтеза аммиака:

→ — технологический поток; — — — → — информационный поток режимно-конструктивных параметров аппаратов; —·—· → — обратные связи, соответствующие проектным условиям;

I—VIII — расчетные модули (в скобках указаны позиции соответствующих аппаратов, изображенных на рис. 11); *IX* — проектное условие $П_с$ = const; *X* — то же, $Г_1$ = const.

целом $П_с$ определяется количеством товарного аммиака, выходящего из танка *13*.

Для моделирования технологической схемы на ЭВМ нужно перейти к ее формализованному математическому описанию. В случае модульного подхода к расчету таким описанием будет расчетная блок-схема и ориентированный граф (рис. 12). Узлы графа представляют модули расчета математических моделей аппаратов, дуги — направления передачи информации от модуля к модулю.

В модуле *I* расчета колонны синтеза использована модель идеализированного реактора, предложенная Ю. А. Соколинским [51]. При этом рассчитывается максимальная производительность реактора, которая теоретически может быть достигнута при данном давлении и составе входящего в колонну газа.

Расчет конденсации аммиака в модулях *II*, *V*, *VIII* требует определения равновесного содержания аммиака a^* в многокомпонентной газовой системе: азот—водород—метан—аргон над жидким аммиаком. По правилу аддитивности

$$a^* = \sum_j a_j^* (P,\ T)\, \bar{c}_j$$

где a^* — равновесное содержание аммиака в системе; $\bar{c}_j$ — содержание j-го компонента в газовой смеси без аммиака, пересчитываемое из содержания j-го компонента c_j в системе по формуле

$$\bar{c}_j = c_j/(1 - a^*)$$

в которой a_j^* $(P,\ T)$ — равновесное содержание аммиака при давлении P и температуре T в бинарной системе аммиак—j-тый компонент, рассчитываемое по уравнению Ларсона—Блэка [52]:

$$a_j^* (P,\ T) = \exp \left(B_{0j} + \frac{B_{1j}}{\sqrt{P}} + \frac{B_{2j}}{\sqrt{T}} \right)$$

Концентрации компонентов в газовой фазе c_j связаны с их растворимостями в жидком аммиаке правилом Генри:

$$v_j = gPc_jH(T)$$

где v_j — количество j-го компонента, растворенного в жидком аммиаке g при давлении P; $H(T)$ — константа Генри, представленная полиномом второй степени от температуры конденсации T.

Количество конденсата аммиака g определяется уравнением материального баланса:

$$\Gamma_{вх}a_{вх} = \Gamma_{вых}a_{вых} + g$$

где Γ_i — расходы; a_i — концентрации аммиака на входе и выходе из аппарата, причем $a_{вых} = a^*$.

При низком давлении в конденсаторе *14* (см. рис. 11) равновесная концентрация аммиака в газовой фазе рассчитывается по закону Дальтона: $a^* = P_{нпNH_3}(T)/P$, где $P_{нпNH_3}$ — давление насыщенных паров аммиака. Структурный анализ системы уравнений баланса конденсатора аммиака показал, что ее решение может быть сведено к итерациям по одной переменной, например, количеству g конденсата аммиака.

Расчет узла смешения *VI* (см. рис. 12) свежего газа с циркуляционным в производстве жидкого аммиака и сборника *VII* несколько более сложен, чем расчет конденсаторов, он включает в себя тепловой баланс смешения потоков, обладающих различным теплосодержанием, при этом происходит испарение или конденсация части жидкого аммиака в зависимости от конкретных условий.

В силу важности производства аммиака его расчету и оптимизации посвящено большое число работ например [53]. Здесь описан расчет отделения синтеза аммиака с помощью автоматизированной системы технологических расчетов (АСТР) [54]. Система АСТР построена по иерархическому принципу и имеет три уровня. На верхнем уровне используются проблемно-ориентированные языки со средствами структурного анализа для автоматического определения порядка расчета: язык СХТС модульного подхода к расчету схемы и язык СОЛВЕК, ориентированный на уравнения [48]; на среднем уровне — ПЛ/1-АСТР — язык ПЛ/1, расширенный специальными синтаксическими и вычислительными средствами ускорения сходимости, оптимизации режимно-конструктивных параметров, печати таблиц материально-тепловых балансов для проектных документов и т. д.; на нижнем уровне — комплексы программ конкретных технологических расчетов, к которым проектировщик обращается с помощью стандартных бланков. Один из таких комплексов — СИНТАМ [54] служит для многовариантных расчетов отделения синтеза аммиака.

Схема отделения синтеза аммиака имеет два технологических цикла (см. рис. 12): цикл синтеза, состоящий из колонны синтеза *I*, первичного конденсатора *II*, линии продувки *III*, вторичного конденсатора *V* со вводом свежего газа *VI*, и рецикл танковых газов, состоящий из танка *VII* и конденсатора танковых газов *VIII*.

В случае проведения проектных расчетов к уравнениям баланса по циклам добавляются уравнения проектных условий, например, такие: подобрать долю отгона продувочных газов α так, чтобы расход циркуляционного газа Γ_1 соответствовал производительности циркуляционного компрессора, или такие: подобрать объем $О_к$ катализатора в колонне синтеза, обеспечивающий требуемую производительность схемы при прочих заданных условиях (см. рис. 12). Информационные обратные связи проектных условий дополняют информационные связи между аппаратами, осуществляемые технологическими потоками; на расчетной схеме (рис. 12) они обозначены штрих-пунктирными линиями. Исходя из технологических соображений в данной схеме можно выделить 11 характеристических величин, т. е. существенных параметров, наиболее часто варьируемых в проектных расчетах. Их можно разбить на два класса.

I. Пять исходных величин (входных переменных схемы) при поверочном (или простом моделирующем) расчете схемы: $\Gamma_{св}$ — расход свежего газа; α — доля отбора продувочных газов; $О_к$ — объем катализатора; $t_{к1}$, $t_{к2}$ — температуры первичной и вторичной конденсации.

II. Шесть результирующих величин поверочного расчета схемы: $П_с$, $П_{кс}$ — производительность схемы и колонны синтеза; a_1, a_2 — содержание аммиака на входе и выходе из колонны синтеза; Γ_1 — расход циркуляционного газа; u_1 — содержание инертов (аргон и метан) в циркуляционном газе.

Виды проектных расчетов одной схемы различаются тем, какие из перечисленных характеристических величин задаются технологом точно, а какие — определяются в ходе итерационного расчета. Таким образом, общее число видов расчетов достигает числа $11!/5!6! = 462$. Возможности блока сходимости ПЛ/1-АСТР обеспечивают объединение всех вариантов расчета в одной программе.

В основном варианте СИНТАМ реализован модульный многоуровневый подход к расчету схем. Первому (внутреннему) уровню итераций сходимости соответствует расчет отдельных аппаратов (в каждом аппарате итерации проводятся не более, чем по двум переменным); второму — расчет рецикла танковых газов по пяти переменным разрываемого потока $Ж_4$ (расход и концентрациям); третьему — баланс цикла синтеза по потоку Γ_1 (пять переменных) и решение уравнений проектных условий (их максимальное число соответствует числу характеристических переменных схемы, относящихся к классу I). Кроме того, использовался модульный двухуровневый подход, при котором итерации по рециклу танковых газов ($Ж_4$) были вынесены на верхний уровень, и подход, ориентированный на уравнения. В последнем случае был проведен структурный анализ всей системы из 166 уравнений материально-теплового баланса отделения синтеза аммиака. Для поверочного расчета общая система разбивается на три блока совместно решаемых уравнений, соответственно, с девятью, двумя и четырьмя итерируемыми переменными. При этом сокращается как число итерируемых переменных (15 против 18 при модульном подходе), так и число итераций сходимости

Таблица 7. **Сравнение методов решения систем нелинейных уравнений**

Метод	Способ выбора шага *	Трехуровневый расчет (по 5 итерируемым переменным на двух верхних уровнях)		Двухуровневый расчет (10 итерируемых переменных на верхнем уровне)	
		число направлений	число обращений к модели	число направлений	число обращений к модели
QNM 1	1	2	20	2	12
QNM 1	2	1	25	2	17
QNM 2	1	13	77	43	149
QNM 2	2	16	307	>50	>400
Broyden 1	1	2	20	2	12
Broyden 1	2	1	25	2	17
Broyden 2	1	24	163	>50	>400
Broyden 2	2	9	166	>50	>400
WOLF	—	3	20	4	13

* 1 — уменьшение нормы вектора невязок; 2 — минимизация нормы вектора невязок по направлению.

(нет вложенных итераций). Время счета уменьшается в среднем в 1,5—2 раза.

Введение проектных условий может по-разному влиять на трудоемкость расчета при модульном и ориентированном на уравнения подходах. В первом случае с каждым проектным условием добавляется новое уравнение, увеличивающее размерность решаемой задачи, во втором эффект может быть противоположным. Так, если заданы производительность схемы Π_c и расход циркуляционного газа Γ_1, общее число итерируемых переменных после структурного анализа сокращается с 15 до 12.

Для решения систем нелинейных уравнений материально-теплового баланса и проектных условий применялись квазиньютоновские методы QNM1, QNM2 (см. с. 47), методы Бройдена Broyden 1, Broyden 2 (см. с. 47) и метод Вольфа (WOLF) [3, с. 35], в который были введены эвристики, касающиеся процедуры построения и изменения симплекса точек, подобранные в ходе многочисленных технологических расчетов различных схем [54].

В табл. 7 приведены результаты сравнения различных методов решения систем уравнений и вариантов модульного подхода к расчету отделения синтеза аммиака. Сравнивались трехуровневый и двухуровневый подходы. В первом случае итерации производились по пяти переменным на верхнем и по пяти переменным на среднем уровне (итерации в итерации). Во втором случае итерации проводились по всем десяти переменным совместно. В обоих случаях итерации на уровне расчета отдельных аппаратов не учитывались. Методы сравнивались по числу направлений одномерного поиска и числу обращений к модели — расчету аппаратов схемы. В случае, когда

число обращений к модели превышало 400, а число направлений — 50 испытание метода прекращалось.

Результаты показывают преимущество совместного решения систем уравнений перед методом вложенных итераций. Незначительный рост числа направлений поиска на верхнем уровне компенсируется значительным сокращением числа обращений к модели. Методы с начальным вычислением обратного Якобиана (QNM1 и Broyden 1) дают лучшие результаты по сравнению с заданием в качестве начального приближения единичной матрицы (QNM2, Broyden 2). Эти последние методы хуже определяют направление поиска корня и, кроме того, оказываются чувствительными к порядку написания уравнений. Использование в одномерном поиске минимизации нормы вектора невязок приводит к увеличению числа обращений к модели, что проявляется особенно заметно, когда направление поиска задано неточно (QNM2, Broyden 2). Метод Вольфа дает в среднем хорошие результаты, хотя это можно отнести за счет заложенных в него эвристик. В целом, можно рекомендовать квазиньютоновские методы с начальным вычислением обратного Якобиана и с уменьшением нормы вектора невязок без минимизации по направлению поиска.

ГЛАВА III

Методы безусловной оптимизации

Общая характеристика задач оптимизации процессов

Рассмотрим алгоритмы оптимизации ХТС, не учитывающие ограничения на варьируемые параметры. Конечно, в любой практической задаче существуют ограничения на эти параметры. Однако на методах безусловной минимизации имеет смысл остановиться по следующим причинам:

а. В задачах оптимизации ХТС ограничения на варьируемые параметры часто имеют простой вид (I, 9). С помощью простой замены переменных задачи с ограничениями этого типа сводятся к задачам без ограничений [55].

б. Методы безусловной минимизации обобщаются (см. гл. IV) на случай учета более сложных линейных ограничений.

в. Решение задач минимизации общего вида часто сводят к последовательности задач безусловной минимизации специальным образом сконструированных функций (см. гл. IV). Таким образом, наличие эффективных алгоритмов решения задач без ограничений

дает основу для решения разнообразных задач оптимизации ХТС при наличии ограничений.

В главе I было показано, что задача оптимизации ХТС эквивалентна задаче оптимизации некоторой функции

$$\min_x f(x) \qquad (III, 1)$$

Остановимся на ряде особенностей вычисления целевой функции $f(x)$ в задачах оптимизации ХТС.

1. Расчет функции $f(x)$ сводится к расчету статического режима ХТС при фиксированных значениях варьируемых параметров.

2. Наличие в схеме аппаратов с распределенными параметрами (каталитические реакторы с неподвижным или «кипящим» слоем катализатора) приводит к необходимости интегрирования систем обыкновенных дифференциальных уравнений при расчете таких аппаратов.

3. В ХТС часто имеются аппараты (реакторы с мешалками, ректификационные колонны, экстракторы и др.), расчет которых сводится к выполнению трудоемких итерационных процедур.

4. Наличие в схеме рециклов и тепловых обратных связей приводит к тому, что при сведении материального и теплового балансов возникает необходимость в итерационной процедуре на уровне расчета ХТС (см. гл. II).

5. Схема может характеризоваться несколькими сотнями фазовых и несколькими десятками управляющих переменных.

Таким образом, задача оптимизации ХТС чрезвычайно сложна, имеет большую размерность, требует огромного числа операций даже для одного расчета значения целевой функции. Следовательно, чтобы решение реальных задач оптимизации ХТС могло быть выполнено в приемлемые сроки необходимо использовать самые эффективные методы оптимизации.

Рассмотрим методы сопряженных направлений и квазиньютоновские методы. В обоих случаях вначале будем рассматривать методы минимизации квадратичных функций, на основе которых будут конструироваться методы минимизации произвольных функций.

Свойства и методы сопряженных направлений

Свойства сопряженных направлений

Пусть требуется найти минимум квадратичной функции

$$f(x) = 0{,}5\,(x,\ Ax) + (b,\ x) + c \qquad (III, 2)$$

где $x = (x_1, \ldots, x_n)$, $b = (b_1, \ldots, b_n)$, $c = (c_1, \ldots, c_n)$ — матрицы-столбцы (векторы) размерности $(n \times 1)$ (b, c — постоянные); $A = (a_{ij})$ — $(n \times n)$-матрица. Будем исходить из предположения, что матрица A:

1. Симметрична

$$A^{\mathrm{T}} = A \qquad (III, 3)$$

2. Положительно-определенна

$$(x,\ Ax) = x^{\mathrm{T}}Ax > 0 \quad \text{для} \quad x \neq 0 \qquad (III, 4)$$

Пусть для поиска минимума используется один из методов спуска, при котором последовательные приближения подсчитываются с помощью формулы (I, 39). Для определения направления p_i на i-той итерации имеется какой-то алгоритм. Коэффициенты α_i на i-той итерации в общем случае выбираются из условия (I, 46), в частном случае, когда ищется минимум на каждом направлении, в точках x_k, ($k = 1, \ldots, n-1$) выполняются соотношения:

$$(p_i,\ g_{i+1}) = 0 \qquad \text{(III, 5)}$$

Градиент g_i функции f в i-той точке будет равен [11, с. 263]

$$g_i = g(x_i) = \operatorname{grad} f(x)\big|_{x=x_i} = Ax_i + b \qquad \text{(III, 6)}$$

Введем обозначение

$$y_i = g_{i+1} - g_i \qquad \text{(III, 7)}$$

Будем исходить из предположения, что $\alpha_i \neq 0$. Используя равенство (III, 6), получим

$$y_i = A(x_{i+1} - x_i) = As_i \qquad \text{(III, 8)}$$

С учетом выражения (I,40) имеем

$$y_i = \alpha_i Ap_i \qquad \text{(III, 9)}$$

Тогда формула (III,7) позволяет получить

$$g_{i+1} = g_i + \alpha_i Ap_i \qquad \text{(III, 10)}$$

Определение. Ненулевые векторы (направления)

$$p_0,\ p_1,\ \ldots,\ p_{n-1} \qquad \text{(III, 11)}$$

будем называть A-сопряженными или просто сопряженными, если выполняются равенства:

$$(p_i,\ Ap_j) = 0 \qquad j = 0,\ 1, \ldots, i-1 \qquad i = 2, \ldots, k \qquad k \leqslant n-1 \qquad \text{(III, 12)}$$

$$(p_i,\ Ap_j) = 0, \qquad i = 0,\ 1, \ldots, j-1 \qquad j = 2, \ldots, k \qquad k \leqslant n-1 \qquad \text{(III, 13)}$$

Поскольку матрица A положительно-определенна, в соответствии с условием (III,4) получим

$$(p_i,\ Ap_i) > 0 \qquad \text{(III, 14)}$$

Равенства (III, 12)—(III.14) будем называть соотношениями сопряженности. Подставив в равенства (III, 12)—(III, 14) выражения $Ap_j = y_j/\alpha_j$ и $p_i = s_i/\alpha_i$ из формул (III, 9), (I, 40), получим другую форму соотношений сопряженности

$$(s_i,\ y_j) = 0 \qquad j = 0,\ 1, \ldots, i-1 \qquad i = 2, \ldots, k \qquad (k \leqslant n-1) \qquad \text{(III, 15)}$$

$$(s_i,\ y_j) = 0 \qquad i = 0,\ 1, \ldots, j-1 \qquad j = 2, \ldots, k \qquad (k \leqslant n-1) \qquad \text{(III, 16)}$$

$$(s_i,\ y_i) > 0 \qquad i = 0,\ 1, \ldots, k \qquad \text{(III, 17)}$$

Введем клеточные $n \times 1$-матрицы Y_i и S_i:

$$Y_i = (y_0, \ldots, y_{i-1}) \qquad S_i = (s_0, \ldots, s_{i-1}) \qquad i \leqslant n \qquad \text{(III, 18)}$$

Используя правила умножения клеточных матриц, соотношения сопряженности (III, 15), (III, 16) можно записать следующим образом:

$$Y_i^{\mathrm{T}} s_i = 0 \qquad i = 1, \ldots, k \qquad k \leqslant n \qquad \text{(III, 19)}$$

$$S_j^{\mathrm{T}} y_j = 0 \qquad j = 1, \ldots, k \qquad k \leqslant n \qquad \text{(III, 20)}$$

Можно показать, что сопряженные векторы (III, 11) являются линейно-независимыми [11].

Конечность числа направлений поиска. Докажем вначале, что в точках смены направлений x_j выполняются соотношения

$$(p_j,\ g_k) = (p_j,\ g_{j+1}) \qquad k \geqslant j+1 \tag{III, 21}$$

Прежде всего ясно, что для $k = j + 1$ соотношение (III,21) верно; докажем, что это справедливо и для $k > j + 1$. Из соотношения (III, 7) для $i = j + 1, \ldots, k$ получим

$$g_k = g_{j+1} + \sum_{i=j+1}^{k-1} y_i \tag{III, 22}$$

Умножим скалярно уравнение (III,22) на p_j $(j + 1 < k)$, тогда согласно свойству (III, 16) сопряженных направлений получим соотношение (III, 21).
Подставляя в соотношение (III,21) при $k = n$ величины g_n и g_{j+1} из равенства (III, 6), легко найдем

$$(p_j,\ Ax_n) = (p_j,\ Ax_{j+1}) \tag{III, 23}$$

Пусть теперь построены n сопряженных направлений $p_0, \ldots, p_{n-1}$ и найдена точка x_n на последнем направлении. Поскольку векторы $p_0, \ldots, p_{n-1}$ являются линейно-независимыми, они образуют базис в n-мерном пространстве [56, с. 72]; отсюда координаты любой точки x этого пространства могут быть представлены в виде:

$$\bar{x} = x_n + \beta_0 p_0 + \cdots + \beta_{n-1} p_{n-1} \tag{III, 24}$$

где β_i — некоторые коэффициенты. Определим теперь коэффициенты β_i таким образом, чтобы точка $\bar{x}$ оказалась минимумом функции f. Для этого координаты точки $\bar{x}$ [с учетом равенства (III, 6)] должны удовлетворять соотношению:

$$\operatorname{grad} f \equiv A\bar{x} + b = 0 \tag{III, 25}$$

Подставим теперь в выражение (III,25) величину $\bar{x}$ из (III, 24) и умножим скалярно это соотношение на вектор p_j; тогда, принимая во внимание свойства (III, 12), (III, 13) сопряженных направлений и используя равенство (III, 23), получим

$$(p_j,\ Ax_{j+1}) + \beta_j (p_j,\ Ap_j) + (p_j,\ b) = 0 \tag{III, 26}$$

Подставляя в (III,26) величину

$$Ax_{j+1} = g_{j+1} - b \tag{III, 27}$$

и используя соотношение (III, 9), легко получим

$$\beta_j = -\alpha_j (p_j,\ g_{j+1})/(p_j,\ y_j) \tag{III, 28}$$

В частном случае, когда точки x_j являются оптимальными на каждом направлении, в соответствии с соотношением (III, 5) все $\beta_j = 0$ $(j = \overline{0,\ n-1})$ и, следовательно, точка x_n является минимумом функции f, т. е. в ней выполняется соотношение

$$g_n = 0 \tag{III, 29}$$

Итак, минимум функции f будет определен за n шагов, если поисковые направления будут сопряженными и на каждом направлении будет определяться оптимальная точка. Если точки смены направлений не будут оптимальными, то потребуется сделать $n + 1$ шаг.

Определение минимума в линейном многообразии. Пусть $p_0, \ldots, p_i$ сопряженные направления. Образуем линейное многообразие M $(x_0,\ p_0, \ldots, p_i)$, проходящее через точку x_0 и параллельное векторам $p_0, \ldots, p_i$. Точки x этого многообразия удовлетворяют соотношениям

$$x = x_0 + \gamma_0 p_0 + \cdots + \gamma_i p_i \tag{III, 30}$$

где γ_l, $(l = \overline{0,\ i})$ произвольные числа. Тогда точка $x_{i+1} = x_i + \alpha_i p_i$ (где x_{i+1} — оптимальная точка на направлении p_i) является минимумом в многообразии M $(x_0,\ p_0, \ldots, p_i)$. Действительно, в силу равенства (III, 21) с учетом соотношения (III, 5) имеем

$$(p_j,\ g_{i+1}) = 0 \qquad j = 0,\ 1,\ \ldots,\ i \tag{III, 31}$$

Из теоремы, приведенной в работе [11, с. 268], следует доказательство сформулированного утверждения.

Построение алгоритмов минимизации. Формы записи соотношений сопряженности (III, 12)—(III, 14) и (III, 15)—(III, 17) совершенно эквивалентны, тем не менее последняя из них имеет то преимущество, что в ней в явном виде не используется матрица A. Это дает возможность строить алгоритмы, в которых не используются коэффициенты матрицы A.

Рассмотрим вначале алгоритмы минимизации квадратичных функций, обеспечивающих выполнение соотношений (III, 15)—(III, 17), у которых на каждом направлении ищется оптимальная точка. В случае симметричности матрицы A выполняется соотношение [II, с. 262]

$$(p_i,\ Ap_j) = (p_j,\ Ap_i) \qquad j < i \qquad \text{(III, 32)}$$

Используя выражения (III, 9), (III, 32), получим

$$(p_i,\ y_j) = \alpha_j\,(p_j,\ Ap_i) = (p_j,\ y_i)\,(\alpha_j/\alpha_i) \qquad j < i \qquad \text{(III, 33)}$$

Выполнение соотношения (III, 15) гарантирует выполнение соотношения (III, 16). Это свойство важно для практического построения алгоритмов минимизации, поскольку последние можно строить таким образом, чтобы они обеспечивали только выполнение соотношения (III, 15).

Перейдем теперь к задаче минимизации неквадратичных функций. Направления, удовлетворяющие условиям (III, 15), будем называть сопряженными [57, с. 123]. Ясно, что в общем случае из соотношений (III, 15) не следуют соотношения (III, 16), как в случае квадратичных функций. Поэтому направления, удовлетворяющие условиям (III,15), не будут обладать свойствами, которыми они обладали бы в случае квадратичности функций, в частности, свойством окончания поиска за n шагов. Тем не менее, алгоритмы, основанные на построении направлений, удовлетворяющих условиям (III,15), оказались эффективными и при минимизации неквадратичных функций. Объясняется это, по-видимому, тем, что в окрестности минимума достаточно гладкая функция должна хорошо аппроксимироваться квадратичной функцией; отсюда алгоритмы, использующие сопряженные направления, должны давать быструю сходимость в окрестности минимума. Заметим, что как раз в окрестности минимума градиентные методы существенно замедляют скорость поиска.

Пусть для произвольной функции построены n направлений $p_0, \ldots, p_{n-1}$, удовлетворяющих условиям (III, 15). Поскольку для произвольной функции оптимальная точка на n-м направлении, вообще говоря, не будет совпадать с точкой минимума целевой функции, необходимо продолжить построение направлений p_n, $p_{n+1}, \ldots$ удовлетворяющих условиям (III, 15).

Рассмотрим построение вектора p_n. В соответствии с условиями (III, 15) он должен быть ортогонален n векторам $y_0, \ldots, y_{n-1}$. Поскольку в общем случае (если векторы $y_0, \ldots, y_{n-1}$ линейно-независимы) в n-мерном пространстве невозможно построить ненулевой вектор, ортогональный n векторам [56, с. 205], при построении вектора p_n можно пойти двумя путями. В первом случае можно отбросить условие $(p_n,\ y_0) = 0$ и потребовать, чтобы выполнялись условия $(p_n,\ y_i) = 0,\ i = 1, \ldots, (n - 1)$. При построении вектора p_{n+1} в свою очередь будет отброшено условие $(p_{n+1},\ y_1) = 0$, и т. д. Тогда на k-м шаге $(k \geqslant n)$ вектор p_k будет строиться таким образом, чтобы векторы $p_k, p_{k-1}, \ldots, p_{k-n+1}$ были сопряженными, т. е. чтобы выполнялись соотношения:

$$(p_k,\ y_{k-1}) = 0 \qquad (p_k,\ y_{k-2}) = 0,\ \ldots,\ (p_k,\ y_{k-n+1}) = 0 \qquad \text{(III, 34)}$$

Во втором случае, начиная с вектора p_n, можно заново предпринять нахождение векторов, удовлетворяющих соотношениям (III,15). Аналогичная ситуация возникнет через n шагов, когда опять надо будет начать новое построение направлений, удовлетворяющих соотношениям (III,15), и т. д. Таким образом, в данном случае при $i = kn$ (где $k = 1, 2, \ldots$) надо будет каждый раз начинать заново строить направления, удовлетворяющие условиям (III, 15).

Будем называть базисом в точке x_k совокупность m векторов $p_k, p_{k-1}, \ldots, p_{k-m+1}$, для которых обеспечивается сопряженность при применении того или иного способа построения сопряженных направлений. Число векторов m будем называть глубиной базиса. Таким образом, при первом способе построения векторов p_i глубина базиса всегда постоянна и равна n. От точки x_k к точке x_{k+1} базис

меняется непрерывно, причем вектор p_{k-m+1} из него удаляется, а вектор p_{k+1} в него добавляется. В связи с этим методы, в которых будет использоваться такой способ построения сопряженных направлений, будем называть методами построения сопряженных направлений с непрерывным изменением базиса.

При втором способе построения сопряженных направлений происходит циклическое изменение базиса. Глубина базиса в каждой точке цикла из n шагов увеличивается на единицу, причем через каждые n шагов построение базиса начинается заново. В связи с этим такой способ построения сопряженных направлений будем называть методом построения сопряженных направлений с циклическим изменением базиса. При использовании подобных методов в точках $i = kn$ $(k = 1, 2, \ldots)$ происходит как бы обновление метода. Подобные алгоритмы называют также алгоритмами с обновлением. Операция обновления может потребоваться и в других случаях. На этом мы остановимся лишь при описании конкретных алгоритмов.

Методы сопряженных направлений *

Рассмотрим методы сопряженных направлений с циклическим и непрерывным изменением базиса.

Метод с циклическим изменением базиса. В соответствии с условиями (III, 15) на i-том шаге $(i < n)$ вектор p_i должен быть ортогонален i векторам $y_0, \ldots, y_{i-1}$, т. е. n компонент вектора p_i удовлетворяют i линейным соотношениям. Это значит, что соотношения (III,15) неоднозначно определяют вектор p_i и имеются $(n - i)$ степеней свободы. В связи с этим можно потребовать, чтобы вектор p_i удовлетворял некоторым дополнительным условиям. Остановимся на одном способе построения p_i. Обозначим через D линейное пространство, натянутое на векторы $y_0, \ldots, y_{i-1}$, а через C его ортогональное дополнение $(C \perp D,\ C \times D = E^n)$. Согласно условиям (III, 15) вектор p_i должен лежать в пространстве C. Помимо этого потребуем, чтобы направление p_i для $i \geqslant 1$ являлось проекцией $-g_i$ на C [31]. В качестве p_0 возьмем $-g_0$. Такой выбор p_i приведет к тому, что угол между антиградиентом и направлением поиска будет наименьшим. Это будет способствовать устойчивости поиска. При таком построении $r_i = p_i/\| p_i \|$ будет направлением наискорейшего убывания функции $f(x)$ в пространстве C, т. е. $r = r_i$ будет давать решение задачи

$$-\max_{\substack{r \\ \|r\|=1}} \frac{df}{dr} \qquad r \in C \tag{III, 35}$$

Поскольку производная df/dr равна $r^{\mathrm{T}}g$, задача (III, 35) нахождения r_i может быть сформулирована следующим образом:

$$\begin{gathered} \min r^{\mathrm{T}}g_i = 0 \qquad i \leqslant n-1 \\ Y_i^{\mathrm{T}}r = 0 \\ r^{\mathrm{T}}r = 1 \end{gathered} \tag{III, 36}$$

Запишем функцию Лагранжа задачи (III, 36)

$$L = r^{\mathrm{T}}g_i + u^{\mathrm{T}}Y_i^{\mathrm{T}}r + \beta\left(r^{\mathrm{T}}r - 1\right)$$

где $u = (u_0, \ldots, u_{i-1})^{\mathrm{T}}$, β — множители Лагранжа. Приравняем нулю частные производные $\partial L/\partial r$ и проделав несложные преобразования, получим [11]

$$p_i = -P_i g_i \tag{III, 37}$$

где

$$P_i = I_n - Y_i\left(Y_i^{\mathrm{T}}Y_i\right)^{-1}Y_i^{\mathrm{T}} \tag{III, 38}$$

Матрицу P_i называют проекционной, так как p_i является проекцией $-g_i$ на C. Из построения и формулы (III, 38) видно, что

$$P_0 = I_n,\ P_n = 0$$

$$\begin{cases} P_i y_j = 0 & \text{для} \quad j = 0, 1, \ldots, i-1 \\ P_i q = q & \text{для таких } q, \text{ что } (q, y_j) = 0 \end{cases}$$

---

* Раздел написан совместно с А. Р. Беляевой.

Покажем, что векторы $y_0, \dots, y_{i-1}$ линейно независимы и, следовательно, существует матрица, обратная $Y_i^{\mathrm{T}} Y_i$. В самом деле, предположим, что $y_0, \dots, y_{i-1}$ линейно зависимы, тогда существует такая совокупность c_j, $(j = \overline{0, i-1})$, $c_0^2 + \dots + c_{i-1}^2 \neq 0$, что

$$c_0 y_0 + c_1 y_1 + \dots + c_{i-1} y_{i-1} = 0 \qquad \text{(III, 39)}$$

Умножим соотношение (III, 39) скалярно на s_j, получим $c_j (y_j, s_j) = 0$. Поскольку $(y_j, s_j) > 0$ [см. выражение (III, 17)], $c_j = 0$ для любого j. Полученное противоречие доказывает линейную независимость $y_0, \dots, y_{i-1}$. В работе [31] показано, что матрицу P_i можно строить с помощью следующего рекуррентного соотношения:

$$P_{i+1} = P_i - \frac{P_i y_i y_i^{\mathrm{T}} P_i}{y_i^{\mathrm{T}} P_i y_i} \quad P_0 = I_n \qquad \text{(III, 40)}$$

Транспонируя выражение для P_{i+1}, легко показать, что, если матрица P_i симметрична, то матрица P_{i+1} также будет симметричной. Но поскольку матрица $P_0 = I_n$ симметрична, все матрицы P_i также будут симметричны. Кроме того, можно показать, что матрица P_i положительно полуопределена. Другой путь построения сопряженных векторов $p_0, \dots, p_i$ основан на использовании процесса ортогонализации Грамма — Шмидта [56, с. 209], [11, с 45]. Рассмотрим теперь конкретные алгоритмы минимизации функций, использующие описанные выше способы построения сопряженных направлений.

Алгоритм I PRM. Оптимальная точка ищется на каждом направлении. Внутри цикла поисковые направления строятся в соответствии с формулами (III, 37), (III, 40). При значениях $i = kn$, $(k = 1, 2, \dots)$ полагается $P_i = I_n$, следовательно, $p_i = -g_i$.

Метод с непрерывным изменением базиса. В этом случае до $i = n - 1$ направления поиска будут определяться решением задачи (III, 36), где матрица Y_i определяется соотношением (III,18), а для $i \geqslant n$ — решением той же задачи, но со следующей матрицей $Y_i = (y_{i-n+1}, \dots, y_{i-1})$.
В этом случае на каждом шаге имеется только одна степень свободы при определении p_i. Решение данной задачи в этом случае также может быть записано в виде формул (III, 37), (III, 38).

Метод сопряженных градиентов [58]. Можно показать, что если оптимальная точка ищется на каждом направлении, то для квадратичных функций метод сопряженных направлений (см. алгоритм I) эквивалентен методу [59, с. 80—81], в котором поисковые направления определяются следующим способом (см. также работу [11, с. 47—49]):

$$p_i = -g_i + \gamma_i p_{i-1} \qquad \text{(III, 41)}$$

Здесь γ_i определяются с помощью одной из следующих формул

$$\gamma_i = (g_i, g_i)/(g_{i-1}, g_{i-1}) \qquad \text{(III, 42)}$$
$$\gamma_i = (y_{i-1}, g_i)/(y_{i-1}, p_{i-1})$$

Алгоритм II основан на методе с циклическим изменением базиса, в котором направления внутри цикла определяются с помощью формул (III, 41), (III,42); обычно его называют методом Флетчера — Ривса.

Преимущество методов сопряженных градиентов по сравнению с методами сопряженных направлений состоит в том, что они требуют хранения только вектора, в то время как методы сопряженных направлений требуют хранения $n \times n$-матрицы. Это особенно важно при решении задач большой размерности. В то же время практический опыт показывает, что методы сопряженных направлений, как правило, обеспечивают бо́льшую скорость сходимости, чем методы сопряженных градиентов. Этот факт, по-видимому, может быть объяснен только тем, что при применении методов сопряженных направлений обеспечивается сопряженность направлений при любых α_i. Метод же сопряженных градиентов этим свойством не обладает, поскольку при выводе формул (III, 41)—(III, 42) существенно использовалось то, что ищется оптимальная точка на каждом направлении [11, с. 47—49].

Минимизация специальных нелинейных функций [60]. Пусть нелинейная функция имеет вид:

$$F = F[f(x)]$$

где $F(f)$ — монотонная функция одной переменной ($F'_f > 0$); $f(x)$ — положительно-определенная квадратичная форма. Используя правило дифференцирования сложных функций, легко получить соотношение

$$\operatorname{grad} f = \frac{1}{F'} \operatorname{grad} F \qquad \text{(III, 42a)}$$

Приравняв grad f нулю, получим, что точка минимума функции F совпадает с точкой минимума квадратичной функции $f(x)$. Используя метод сопряженных градиентов (III, 41), (III, 42) для минимизации функции $f(x)$, найдем ее минимум, а, следовательно, и минимум функции $F\ [f(x)]$, за n шагов. Будем считать, что во время поиска мы можем получать значения только функции F и ее градиента. В связи с этим преобразуем формулу (III, 41) таким образом, чтобы в нее входили только упомянутые значения. Для этого подставим в формулу (III, 41) значение grad f из соотношения (III, 42a):

$$p_i = -\frac{1}{F'_i} \operatorname{grad} F_i + \frac{(F'_{i-1})^2}{(F'_i)^2} \frac{(\operatorname{grad} F_i,\ \operatorname{grad} F_i)}{(\operatorname{grad} F_{i-1},\ \operatorname{grad} F_{i-1})} p_{i-1} \qquad \text{(III, 43)}$$

где через F'_i обозначено значение F'_f в i-той точке. Вводим векторы $\bar{p}_i = F'_i p_i$; поскольку p_i и $\bar{p}_i$ коллинеарны и ищется минимум на направлении, безразлично, какой из этих векторов использовать. Подставляя значение $\bar{p}_i$ в выражение (III, 43), после простых преобразований получим

$$\bar{p}_i = -\operatorname{grad} F_i + \rho_i \frac{\|\operatorname{grad} F_i\|^2}{\|\operatorname{grad} F_{i-1}\|^2} \bar{p}_{i-1} \qquad \rho_i = \frac{F'_{i-1}}{F'_i}$$

Далее, предполагая тот или иной вид функции $F(f)$, можно строить различные конкретные формулы. Так, в работах [60; 61] предлагаются формулы, основанные на представлении F в виде $F = \varepsilon_1 f(x) + 0{,}5\varepsilon_2 f^2(x)$, $F = \varepsilon_1\ [\exp(f(x) - 1)]$. Коэффициенты ε_1, ε_2 неизвестны и подбираются на основе данных, получаемых во время итераций. Для ряда тестовых функций показано превосходство этого подхода перед обычным методом сопряженных градиентов [60; 61].

Квазиньютоновские методы оптимизации

Метод Ньютона, обеспечивающий минимизацию произвольных функций, описан в работе [11, с. 268]. Основным недостатком этого метода является необходимость на каждом шаге вычислять матрицу вторых производных (гессиан) функции $f(x)$. Это обстоятельство явилось побудительной причиной развития квазиньютоновских методов, в которых на основе информации о значениях функции и ее производных в точках поиска строится некоторая аппроксимация либо самого гессиана B_i, либо обратного гессиана H_i (i — номер точки).

Ясно, что задача поиска минимума произвольной достаточно гладкой функции может быть сведена к решению системы нелинейных уравнений

$$\operatorname{grad} f = 0 \qquad \text{(III, 44)}$$

дающих необходимые условия минимума функции $f(x)$. Поскольку матрица Якоби J системы (III, 44) является гессианом G функции $f(x)$, все соотношения (II, 14), (II, 28), (II, 30) будут выполняться, только вместо J в них должна стоять матрица G, а вектор y_i будет

определяться формулой (III, 7). Аппроксимация B_i (самого гессиана) и H_i (обратного гессиана) должны удовлетворять соотношениям (II, 25), (II, 29), (II, 33) и соотношениям (II, 31), (II, 32), (II, 34) соответственно.

Для решения системы (III, 44) может быть применен любой из квазиньютоновских методов, описанных в главе II. Тем не менее, имеет смысл самостоятельно рассмотреть задачу разработки квазиньютоновских методов минимизации, поскольку имеется ряд свойств этой задачи, использование которых позволяет строить более эффективные алгоритмы. Остановимся на этих свойствах.

1. Гессиан G является симметричной матрицей

$$G^{\text{т}} = G \tag{III, 45}$$

2. В окрестности минимума функция $f(x)$ может быть достаточно хорошо аппроксимирована квадратичной формой с положительно определенной матрицей. Это значит, что в достаточно малой окрестности точки минимума функции f гессиан G будет положительно определенным

$$G > 0 \tag{III, 46}$$

3. В каждой точке x_i известно направление (это направление антиградиента), которое дает наискорейшее уменьшение функции

В случае решения систем нелинейных уравнений аналогов этих свойств нет. При построении приближений к гессиану (обратному гессиану) естественно потребовать, чтобы они удовлетворяли возможно большему числу свойств, которыми обладает сам гессиан (обратный гессиан). Имея в виду свойства (III, 45), в большинстве случаев будет требовать, чтобы матрицы B_i, H_i были симметричными

$$B_i^{\text{т}} = B_i \qquad H_i^{\text{т}} = H_i \tag{III, 47}$$

Кроме того, желательно, чтобы матрицы B_i, H_i, были положительно определенными

$$B_i > 0, \; H_i > 0 \tag{III, 48}$$

После определения матрицы H_i (B_i) направление поиска p_i [по аналогии с ньютоновским шагом, см. выражения (II, 12), (II, 13)] находится либо с помощью уравнения (I, 41), либо решением системы линейных уравнений

$$B_i p_i = -g_i \tag{III, 49}$$

Свойство 3 обычно используется для проверки того, является ли данное направление p_i направлением спуска. Поскольку выполняется формула

$$\frac{df}{dp_i} = \frac{p^{\text{т}}}{\| p_i \|} g_i$$

проверяют условия

$$p_i^{\text{т}}(-g_i) > 0 \tag{III, 50}$$

При их выполнении производная df/dp меньше нуля, и направление p_i есть направление спуска. Если выполняется условие (III, 48), то неравенство (III, 50) удовлетворяется автоматически [см. выражение (I, 44)].

Аналогично тому, как было сделано при разработке квазиньютоновских методов решения систем нелинейных уравнений, рассмотрим квазиньютоновские методы 1-го рода, в которых матрицы B_i, H_i будут удовлетворять соотношениям (II, 25,) (II, 31) соответственно, и квазиньютоновские методы 2-го рода, в которых матрицы B_i, H_i будут удовлетворять соотношениям (II, 29), (II, 32) соответственно. Так же, как и в главе II воспользуемся соотношениями (II, 38), (II, 56).

Для простоты обозначений вместо матрицы B_{i+1}, H_{i+1} и B_i, H_i будем иногда писать $\overline{B}$, $\overline{H}$ и B, H, соответственно. Для построения квазиньютоновских методов 1-го рода используем вариационный подход, а для построения квазиньютоновских методов 2-го рода — технику псевдообратных матриц. При применении вариационного подхода необходим критерий, минимизация которого дает возможность определить матрицы E, D. Прежде всего, конечно, применим принцип наименьшего изменения аппроксимирующих матриц на каждой итерации, при котором в качестве критерия используется норма Фробениуса матриц E или D. Основное отличие вывода квазиньютоновских методов минимизации 1-го рода от выхода квазиньютоновских методов 1-го рода, предназначенных для решения систем нелинейных уравнений, будет состоять в требовании симметричности матриц B_i, H_i, т. е. в выполнении условий (III, 47).

Свойство 3 дает возможность поставить другую экстремальную задачу для получения матрицы E. А именно, матрица E может быть найдена из условия минимизации угла между направлением поиска и направлением градиента. В этом случае решается задача

$$\min p_{i+1}^{\mathrm{T}} g_{i+1} \qquad \| p_i \| = 1 \qquad \text{(III, 51)}$$

при выполнении квазиньютоновских и каких-либо других условий. Другими словами в этом случае ищется направление наискорейшего спуска при выполнении квазиньютоновских и других условий.

Квазиньютоновские методы минимизации 1-го рода

Вначале рассмотрим случай построения приближения к самому гессиану. Для получения матрицы E [см. равенство (II, 38)] воспользуемся принципом наименьшего изменения матрицы B_j. При этом матрица B_{j+1} должна удовлетворять квазиньютоновскому условию 1-го рода (II, 25) и условию симметричности (III, 47). Отсюда следует, что матрица E должна удовлетворять условию (II, 39) и условию симметричности. Итак, матрица E будет определяться как решение следующей задачи

$$\min_{e_{ij}} \| E \|_F \qquad \text{(III, 52)}$$

$$E s_j = r_j \qquad \text{(III, 53)}$$

$$E^{\mathrm{T}} = E \qquad \text{(III, 54)}$$

где

$$r_j = y_j - B_j s_j \qquad \text{(III, 55)}$$

Чтобы удовлетворить условию симметричности (III, 54) будем искать матрицу E в виде:

$$E = 0{,}5\,(C + C^{\mathrm{T}}) \qquad \text{(III, 56)}$$

где C — произвольная матрица. Задача (III, 52)—(III, 54) перепишется в виде

$$\min_{c_{ij}} \| C + C^{\mathrm{T}} \|_F \qquad (C + C^{\mathrm{T}})\, s_j = 2r_j$$

Для решения этой задачи воспользуемся методом множителей Лагранжа. Функция Лагранжа будет иметь вид:

$$L = \frac{1}{8} \sum_{k=1}^{n} \sum_{l=1}^{n} (c_{kl} + c_{lk})^2 + \sum_{k=1}^{n} \lambda_k \sum_{l=1}^{n} (c_{kl} + c_{lk})\, s_{l,\,k}$$

где c_{lk} — элемент матрицы C. Приравняв производные L по c_{ip} нулю, получим

$$0{,}25\,(c_{ip} + c_{pi}) + 0{,}25\,(c_{pi} + c_{ip}) + \lambda_i s_{p,\,j} + \lambda_p s_{i,\,j} = 0$$

Из равенства (III, 56) вытекает, что $e_{ip} = 0{,}5\,(c_{ip} + c_{pi})$ предыдущее равенство принимает вид

$$e_{ip} = -\,(\lambda_i s_{p,\,j} + \lambda_p s_{i,\,j}) \qquad \text{(III, 57)}$$

Легко проверить, что соотношения (III, 57) эквивалентны следующему матричному соотношению:

$$E = -\left(\lambda s_j^{\mathrm{T}} + s_j \lambda^{\mathrm{T}}\right) \qquad \text{(III, 58)}$$

Представляя это значение E в уравнение (III, 53), получим:

$$-\lambda s_j^{\mathrm{T}} s_j - s_j \lambda^{\mathrm{T}} s_j = r_j$$

Выразим λ через остальные члены

$$\lambda = \frac{1}{\alpha}\,(-r_j - \beta s_j) \qquad \text{(III, 59)}$$

где $\alpha = s_j^{\mathrm{T}} s_j$; $\beta = \lambda^{\mathrm{T}} s_j$.

Заметим, что β — неизвестная константа, поскольку она зависит от величины λ. Для определения β подставим выражение (III, 59) для λ в уравнение (III, 58), и после несложных преобразований получим

$$\left(-2\beta_j + \frac{1}{\alpha}\, r_j^{\mathrm{T}} s_j\right) s_j = 0 \qquad \text{(III, 60)}$$

Однако вектор s_j не равен нулю; поэтому сумма, стоящая в скобках в равенстве (III, 60), равна нулю, следовательно,

$$\beta = -r_j^{\mathrm{T}} s_j / 2\alpha \qquad \text{(III, 61)}$$

Подставляя значение λ из выражения (III, 58) в формулу (III, 57) и используя значение β из выражения (III, 61), получим

$$E = \frac{1}{\alpha}\left(r_j s_j^{\mathrm{T}} + s_j r_j^{\mathrm{T}}\right) - \frac{1}{\alpha^2}\, r_j^{\mathrm{T}} s_j s_j s_j^{\mathrm{T}} \qquad \text{(III, 62)}$$

Воспользовавшись выражением (III, 55) для r_j и соотношением (II, 38), получим рекуррентную формулу для определения B_i

$$B_{j+1} = B_j + \frac{(y_j - B_j s_j) s_j^{\mathrm{T}} + s_j (y_j - B_j s_j)^{\mathrm{T}}}{s_j^{\mathrm{T}} s_j} - \frac{(y_j - B_j s_j)^{\mathrm{T}} s_j s_j s_j^{\mathrm{T}}}{(s_j^{\mathrm{T}} s_j)^2} \quad \text{(III, 63)}$$

Эта формула была впервые получена Пауэллом, который применил к формуле Бройдена (II, 47) технику симметризации [62, с. 31]. Рекуррентная формула носит название преобразование PSB (Powell—symmetric—Broyden), а матрица B_{j+1}, подсчитываемая с помощью формулы (III, 63), обозначается через $\bar{B}_{PSB}$. Остановимся теперь на случае, когда ищется аппроксимация H_i обратного гессиана. Матрица H_{j+1} ищется в виде (II, 56). Для определения D_j опять используется принцип наименьшего изменения матрицы при условии выполнения квазиньютоновских соотношений (II, 31) [см. также (II, 58)] и условия симметричности (III, 47). Итак, задача определения будет иметь вид

$$\min_{d_{ij}} \| D \|_F$$

$$Dy_j = m_j \qquad D^{\mathrm{T}} = D \quad \text{(III, 64)}$$

где $m_j = s_j - H_j y_j$.

Действуя аналогично выводу формулы (III, 63), можно получить [27]

$$H_{j+1} = H_j + D_{Gr}$$

где

$$D_{Gr} = \frac{(s_j - H_j y_j) y_j^{\mathrm{T}} + y_j (s_j - H_j y_j)^{\mathrm{T}}}{y_j^{\mathrm{T}} y_j} + \frac{(s_j - H_j y_j)^{\mathrm{T}} y_j y_j y_j^{\mathrm{T}}}{(y_j^{\mathrm{T}} y_j)^2} \quad \text{(III, 65)}$$

Получим теперь выражение для B_{j+1}, исходя из условия, что в качестве критерия минимизации при определении матрицы E использована норма Фробениуса некоторой взвешенной матрицы

$$V = WEW \quad \text{(III, 66)}$$

где W — симметричная, невырожденная $n \times n$-матрица [28]. Из формулы (III, 66) имеем

$$E = W^{-1} V W^{-1} \quad \text{(III, 67)}$$

Итак, матрица E определяется как решение задачи

$$\min_{e_{ij}} \| WEW \|_F \quad \text{(III, 68)}$$

$$Es_j = r_j, \quad E^{\mathrm{T}} = E$$

Подставляя выражение (III, 67) для E в (III, 68), придем к следующей задаче

$$\min_{v_{ij}} \| V \|_F \quad \text{(III, 69)}$$

$$V \bar{s}_j = \bar{r}_j \qquad V^{\mathrm{T}} = V \quad \text{(III, 70)}$$

где

$$\bar{s}_j = W^{-1} s_j; \ \bar{r}_j = W y_j - W B_j s_j \quad \text{(III, 71)}$$

Задача (III, 69) выглядит так же, как задача (III, 52)—(III, 54), только роль матрицы E играет матрица V, а векторы s_j, r_j заменены

соответственно векторами $\bar{s}_j$, $\bar{r}_j$. По аналогии с (III, 62) мы можем сразу выписать выражение для V

$$V = \frac{\left(\bar{r}_j \bar{s}_j^{\mathrm{T}} + \bar{s}_j \bar{r}_j^{\mathrm{T}}\right)}{\bar{s}_j^{\mathrm{T}} \bar{s}_j} - \frac{\bar{r}_j^{\mathrm{T}} \bar{s}_j^{\mathrm{T}} \bar{s}_j \bar{s}_j^{\mathrm{T}}}{\left(\bar{s}_j^{\mathrm{T}} \bar{s}_j\right)^2} \qquad \text{(III, 72)}$$

Подставляя вместо r_j, s_j их значения из формул (III, 71), умножая полученное выражение слева на W^{-1} и справа на W и используя формулу (III, 67), получим выражение для E:

$$E = \frac{(y_j - B_j s_j)(M s_j)^{\mathrm{T}} + (M s_j)(y_j - B_j s_j)}{(s_j,\ M s_j)} - \frac{s_j^{\mathrm{T}} (y_j - B_j s_j)(M s_j)(M s_j)^{\mathrm{T}}}{(s_j,\ M s_j)^2} \qquad \text{(III, 73)}$$

где

$$M = W^{-2} \qquad \text{(III, 74)}$$

Введем обозначение $c = Ms$, подставим выражение (III, 73) для E в формулу (II, 38) и используя для B_{i+1}, H_{i+1} обозначения $\overline{B}$, $\overline{H}$, соответственно, получим

$$\overline{B} = B_j + \frac{(y_i - B_j s_j)\, c^{\mathrm{T}} + c\,(y_j - B_j s_j)^{\mathrm{T}}}{(s_j,\ c)} - \frac{s_j^{\mathrm{T}} (y_j - B_j s_j)\, c c^{\mathrm{T}}}{(s_j,\ c)^2} \qquad \text{(III, 75)}$$

Получим теперь выражение для H_j, исходя из условия, что в качестве критерия минимизации при определении матрицы D будет использована норма Фробениуса некоторой взвешенной матрицы WDW (где W — симметричная, невырожденная $n \times n$-матрица). В данном случае матрица D будет определяться как решение задачи

$$\min_{d_{ij}} \| WDW \|_F \qquad \text{(III, 76)}$$

$$D y_j = m_j \qquad D^{\mathrm{T}} = D$$

Проводя вывод такой же, как и в предыдущем случае, легко получить

$$\overline{H} = H_j + \frac{(s_j - H_j y_j)\, b^{\mathrm{T}} + b\,(s_j - H_j y_j)^{\mathrm{T}}}{(y_j,\ b)} - \frac{y_j^{\mathrm{T}} (s_j - H_j y_j)\, b b^{\mathrm{T}}}{(y_j,\ b)^2} \qquad \text{(III, 77)}$$

где

$$b = Q y_j,\ Q = W^{-2} \qquad \text{(III, 78)}$$

В формулах (III, 75), (III, 77) имеются произвольные матрицы M, Q, конкретный выбор которых позволяет получить различные формулы для определения $\overline{B}$, $\overline{H}$. Если матрицу M в выражении (III, 75) выбрать так, чтобы она удовлетворяла уравнению

$$M s_j = y_j \qquad \text{(III, 79)}$$

то полученная матрица $\overline{B}$ будет обладать следующим свойством: обратная ей матрица $\overline{B}^{-1}$ будет давать известную формулу Давидона—Флетчера—Пауэлла (DFP) [63]. В связи с этим обозначим матрицу $\overline{B}$ через $\overline{B}_{DFP}$, а обратную ей — через $\overline{H}_{DFP}$; тогда из соотношения (III, 75) имеем

$$\overline{B}_{DFP} = B_j + \frac{(y_j - B_j s_j)\, y_j^{\mathrm{T}} + y_j\,(y_j - B_j s_j)^{\mathrm{T}}}{(s_j,\ y_j)} - \frac{s_j^{\mathrm{T}} (y_j - B_j s_j)\, y_j y_j^{\mathrm{T}}}{(s_j,\ y_j)^2} \qquad \text{(III, 80)}$$

а $\overline{H}_{DFP}$ имеет вид

$$\overline{H}_{DFP} = H_j + \frac{s_j s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} y_j} - \frac{H_j y_j y_j^{\mathrm{T}} H_j}{y_j^{\mathrm{T}} H_j y_j} \tag{III, 81}$$

Доказательство факта, что $\overline{H}_{DFP} = \overline{B}_{DFP}^{-1}$ проводится непосредственной проверкой соотношения $\overline{H}_{DFP} \cdot \overline{B}_{DFP} = I_n$. Если матрицу Q в формуле (III, 77) выбрать удовлетворяющей уравнению

$$Q y_j = s_j \tag{III, 82}$$

то придем к формуле

$$\overline{H} = H_j + \frac{(s_j - H_j y_j) s_j^{\mathrm{T}} + s_j (s_j - H_j y_j)^{\mathrm{T}}}{(y_j,\ s_j)} - \frac{y_j^{\mathrm{T}} (s_j - H_j y_j) s_j s_j^{\mathrm{T}}}{(y_j,\ s_j)^2} \tag{III, 83}$$

которая после несложных преобразований может быть приведена к виду, представляющему известную формулу Бройдена—Флетчера—Гольдфарба—Шенно (*BFGS*) [64; 65]:

$$\overline{H}_{BFGS} = H_j + \left[1 + \frac{y_j^{\mathrm{T}} H_j y_j}{s_j^{\mathrm{T}} y_j}\right] \frac{s_j^{\mathrm{T}} s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} y_j} - \frac{s_j y_j^{\mathrm{T}} H_j + H_j y_j s_j^{\mathrm{T}}}{s_j^{\mathrm{T}} y_j} \tag{III, 84}$$

Проанализируем полученные преобразования (III, 63), (III, 65), (III, 81), (III, 83). Вычислительный опыт показывает [27], что формула (III, 65) работает не вполне удовлетворительно, лучшие результаты дает формула *PSB* (III, 63). Однако и это преобразование уступает формулам (III, 80), (III, 83). Формула (III, 81) (метод *DFP*) получила очень большое распространение и во многих случаях показывает очень большую эффективность. И наконец, преобразование *BFGS* (III, 83) считается одним из лучших с точки зрения быстроты сходимости.

Эти формулы будут получены также при рассмотрении квазиньютоновских методов 2-го рода. Там же будет показано что несмотря на то,что в обоих случаях матрица B_i, H_i удовлетворяет только одному квазиньютоновскому условию (II, 25), (II, 31) в текущей точке поиска, оба эти метода обеспечивают нахождение минимума положительно определенной квадратичной формы за конечное число шагов. Таким образом, если при конструировании квазиньютоновских методов расчета систем нелинейных уравнений лучшие результаты дает минимизация нормы Фробениуса матрицы E при аппроксимации самой матрицы Якоби, то при конструировании квазиньютоновских методов минимизации лучшие результаты дает минимизация нормы Фробениуса взвешенной матрицы E при аппроксимации самого гессиана и минимизация взвешенной матрицы D при аппроксимации обратного гессиана.

Попробуем дать этому факту качественное объяснение [28]. Будем исходить из предположения, что M положительно определенная матрица, тогда существует матрица $M^{1/2}$. С учетом (III, 74) соотношение (III, 66) может быть переписано в виде

$$V = M^{-1/2} E M^{-1/2} \tag{III, 85}$$

Произведем замену переменных $\bar{x} = M^{1/2}x$ для функции $f(x)$, тогда $\bar{f}(\bar{x}) = f(M^{-1/2}\bar{x})$. Гессиан функции $f(x)$ равен

$$\nabla^2 \bar{f}_{\bar{x}\bar{x}} = M^{-1/2}\nabla^2 f M^{-1/2} \qquad (III, 86)$$

Отсюда матрицы

$$M^{-1/2}BM^{-1/2} \text{ и } M^{-1/2}\overline{B}M^{-1/2} \qquad (III, 87)$$

будут давать приближение к гессиану функции $\bar{f}(x)$ в точках $\bar{x}_i$, x_{i+1} соответственно. Поставим теперь задачу построить квазиньютоновский метод 1-го рода для минимизации функции $f(\bar{x})$. Используя принцип наименьшего изменения аппроксимирующей матрицы, по аналогии с задачей (III, 51)—(III, 53) получим

$$\min \| M^{-1/2}\overline{B}M^{-1/2} - M^{-1/2}BM^{-1/2} \|_F \qquad (III, 88)$$

$$M^{-1/2}\overline{B}M^{-1/2}\bar{s}_j = \bar{y}_j \qquad (III, 89)$$

$$(M^{-1/2}\overline{B}M^{-1/2})^{\tau} = (M^{-1/2}\overline{B}M^{-1/2}) \qquad (III, 90)$$

где через $\bar{s}_j$, $\bar{y}_j$ обозначены величины s_j, y_j в пространстве переменных $\bar{x}$.

Покажем, что задача (III, 88)—(III, 90) сводится к задаче (III, 68). Действительно, подставив выражение (II, 38) в выражение (III, 88) и используя формулу (III, 74), получим выражение (III, 68). Используя симметричность матрицы M из (III, 90), легко получить условие $\overline{B}^{\tau} = \overline{B}$ и отсюда условие $E^{\tau} = E$. Далее заметим, что имеют место соотношения $\bar{s}_j = M^{1/2}s_j$, $\bar{y}_j = M^{1/2}y_j$. Подставляя эти выражения в (III, 89) после несложных преобразований, получим условие $\overline{B}s_j = y_j$, которое эквивалентно условию $Es_j = r_j$ (с. 34). Итак, задача (III, 68) дает матрицу E, следовательно и $\overline{B}$, на основе которой мы получаем квазиньютоновские преобразования (III, 75) для пространства $\bar{x}$.

Почему же все-таки преобразование, полученное на основе задачи (III, 68) обеспечивает хорошие практические результаты. Дело в том, что пространство $\bar{x}$ можно сделать очень удобным для поисковых методов соответствующим подбором матрицы M. Действительно, если выбрать $M = \nabla^2 f(x^*)$, где x^* — минимум функции $f(x)$, то в точке минимума гессиан (III, 86) функции $\bar{f}(\bar{x})$ будет равен единичной матрице. Таким образом, в пространстве $\bar{x}$ в окрестности минимума поверхности уровня функции $\bar{f}(\bar{x})$ близки к гиперсферам. Это наиболее благоприятная ситуация для поисковых методов, и если бы мы действительно могли выбрать $M = \nabla^2 f(x^*)$, это объяснило бы причины хорошей работы преобразования (III, 75). К сожалению, мы заранее не знаем матрицы $\nabla^2 f(x^*)$. Однако можно считать, что матрица M является некоторым приближением к матрице $\nabla^2 f(x^*)$, поскольку она является симметричной, положительно определенной и удовлетворяет квазиньютоновскому условию (III, 79). Отсюда и преобразование (III, 75), полученное с использованием этой матрицы, должно показывать результаты, близкие к тем, которые бы мы имели при $M = \nabla^2 f(x^*)$.

Рассмотрим теперь случай, когда ищется аппроксимация к обратному гессиану. Будем предполагать, что матрица Q [см. выражения (III, 78)] является положительно определенной.

Введем замену переменных $\bar{x} = Q^{-1/2}x$ и проанализируем функцию $\bar{f}(\bar{x}) = f(Q^{1/2}\bar{x})$. Гессиан и обратный гессиан функции $\bar{f}(\bar{x})$ равны $Q^{1/2}\nabla^2 f Q^{1/2}$, $Q^{-1/2}(\nabla^2 f)^{-1}Q^{-1/2}$, соответственно. Приближение к обратному гессиану функции $\bar{f}(\bar{x})$ будет иметь вид $Q^{-1/2}HQ^{-1/2}$, где по-прежнему H есть приближение к обратному гессиану функции $f(x)$. Проведя рассуждение такие же, как и в предыдущем случае, можно показать, что задача (III, 76) эквивалентна построению приближения к обратному гессиану в пространстве $\bar{x}$. Отсюда, если матрица Q будет близка к обратному гессиану в точке минимума, то в окрестности точки минимума функция $\bar{f}(\bar{x})$ будет иметь гессиан, близкий к единичной матрице, и, следовательно, мы опять приходим к ситуации, наиболее благоприятной для методов спуска. Но функция Q является симметричной, положительно определенной и удовлетворяет квазиньютоновскому условию (III, 82), поэтому можно считать, что она действительно является приближением к обратному гессиану. Это и объясняет хорошую работу формулы (III, 84).

Проведенные рассуждения имеют значение не только с точки зрения объяснения хорошей работы формул (III, 81), (III, 84), они обосновывают целесообразность использования взвешенных матриц WEW, WDW при рассмотрении новых случаев.

Квазиньютоновские методы минимизации 2-го рода

Вначале рассмотрим методы минимизации квадратичных функций (III, 2). Аппроксимация B_j прямого и H_j обратного гессианов будем искать с помощью соотношений (II, 29) и (II, 32) соответственно. На первой стадии будем вести изложение применительно к случаю, когда ищется матрица H_i. В статье [33] (см. также [11, с. 63]) показано, что, если векторы p_i строятся в соответствие с формулой (I,41), в которой H_i удовлетворяет соотношению (II, 32), а коэффициенты α_i — условию (I, 47), то поисковые направления будут сопряженными, и, следовательно, минимум функции (III, 2) будет найден на n-м шаге. Кроме того, в n-й точке будет выполняться соотношение

$$H_n = A^{-1} \tag{III, 91}$$

Действительно, используя формулу (III, 8) и правило перемножения клеточных матриц, легко проверить, что матрицы Y_j, S_j связаны следующими соотношениями:

$$Y_j = AS_j \qquad S_j = A^{-1}Y_j \qquad Y_j^{\mathrm{T}} = S_j^{\mathrm{T}}A \tag{III, 92}$$

Пусть теперь построены n ненулевых сопряженных векторов s_0, ..., s_{n-1}. Положим $i = n$ в соотношении (II, 32), тогда

$$H_nY_n = S_n \tag{III, 93}$$

Как было показано (см. с. 85), векторы y_0, ..., y_{n-1} в этом случае линейно-независимы. Следовательно, существует обратная мат-

рица Y_n^{-1}. Умножая равенство (III, 93) справа на матрицу Y_n^{-1} и подставляя в полученное выражение S_n из соотношений (III, 92), легко получить формулу (III, 91).

Определение вида матрицы H_i. В гл. II было получено общее решение системы (II, 32). Оно имеет вид (II, 90), (II, 91). Это общее решение можно применить и в данном случае. Могут использоваться также и все частные формулы (II, 97)—(II, 102), которые были получены из общего выражения при некоторых частных значениях произвольных постоянных.

Остановимся еще раз на проблеме определения вектора c_i в формуле (II, 101); как уже указывалось (см. с. 44) вектор c_i может быть выбран:

1) как столбец единичной матрицы I_n;

2) из условия максимума абсолютной величины знаменателя правой части формулы (II, 101);

3) из условия максимума абсолютной величины определителя матрицы H_{i+1};

4) из условия минимума нормы Фробениуса второго члена правой части формулы (II, 101).

Свойство 3 оптимизационной задачи (см. с. 87) позволяет находить c_i из условия минимальности (при определенных условиях) угла между направлениями антиградиента и поиска. Математически задача определения вектора c_i в данном случае формулируется в виде (III, 51), где p_{i+1} определяется формулой (I, 41), а H_{i+1} — соотношением (II, 101). Однако условие $\|p_{i+1}\| = 1$ приводит к существенному усложнению задачи. Поэтому заменим его более простым: $\|c_i\| = 1$. Подставляя в выражение (III, 51) значение p_{i+1} из формулы (I,41) и H_{i+1} из формулы (II, 101) и заменяя условие $\|p_{i+1}\| = 1$ условием $\|c_i\| = 1$, придем к следующей задаче:

$$\min_c g_{i+1}^{\mathrm{T}} \frac{(s_j - H_j y_j)\, c_j^{\mathrm{T}} K_i}{c^{\mathrm{T}}{}_i K_i y_i}\, g_{i+1} \qquad \| c \|^2 = 1 \qquad \text{(III, 94)}$$

Поскольку $g_{i-1}^{\mathrm{T}}(s_i - H_i y_i)$ есть скаляр, а $c_i^{\mathrm{T}} K_i$, $K_i y_i$ — векторы, это выражение является задачей типа (II, 77), решение которой было уже рассмотрено.

Заметим, что матрица H_{i+1}, определяемая формулами (II,90), (II,91), не обладает ни свойством симметричности, ни свойством положительной определенности. Поэтому желательно получить частные формулы этого семейства, которые обладали бы этими свойствами. Покажем вначале, что общая формула

$$H_{i+1} = H_i + \frac{s_i\left(\alpha_{1i} y_i^{\mathrm{T}} H_i + \beta_{1i} s_i^{\mathrm{T}}\right)}{\left(\alpha_{1i} y_i^{\mathrm{T}} H_i + \beta_{1i} s_i^{\mathrm{T}}\right) y_i} - H_i \frac{y_i\left(\gamma_{1i} y_i^{\mathrm{T}} H_i + \delta_{1i} s_i^{\mathrm{T}}\right)}{\left(\gamma_{1i} y_i^{\mathrm{T}} H_i + \delta_{1i} s_i^{\mathrm{T}}\right) y_i} \qquad \text{(III, 95)}$$

полученная в работе [66], является частным случаем формул (II,90), (II,91). Доказательство будет проведено по индукции.

Пусть в точке i выполняются условия

$$s_i^{\mathrm{T}} Y_i = 0 \qquad \text{(III, 96)}$$

$$H_i Y_i = S_i \qquad \text{(III, 97)}$$

Тогда в точке $i+1$ будут выполняться условия

$$s_{i+1}^{\mathrm{T}} Y_{i+1} = 0 \tag{III, 98}$$

$$H_{i+1} Y_{i+1} = S_{i+1} \tag{III, 99}$$

Равенство (III,96) является условием сопряженности [см. выражение (III,19)]. Соотношение (III,98) доказано (см. [11, с. 63]), поэтому докажем правильность соотношения (III,99). Пусть постоянные d_{ji}, c_{ji} $(j = 1, 2)$ определяются с помощью формул:

$$d_{ji} = \alpha_{ji} H_i^{\mathrm{T}} y_i + \beta_{ji} s_i \qquad j = 1, 2 \tag{III, 100}$$

$$c_{ji} = \gamma_{ji} H_i^{\mathrm{T}} y_i + \delta_{ji} s_i \qquad j = 1, 2 \tag{III, 101}$$

где α_{ji}, β_{ji}, γ_{ji}, δ_{ji} — произвольные постоянные. Используя выражение (III,100) для d_{ji} и (II,88) для K_i^j, получим:

$$d_{ji}^{\mathrm{T}} K_i^j = \alpha_{ji} y_i^{\mathrm{T}} H_i + \beta_{ji} s_i^{\mathrm{T}} - \alpha_{ji} y_i^{\mathrm{T}} H_i Y_i Y_i^j - \beta_{ji} s_i^{\mathrm{T}} Y_i Y_i^j \tag{III, 102}$$

На основании равенства (III,96) последний член выражения (III,102) равен нулю. Используя соотношения (III,97), (III,20), легко показать, что предпоследний член правой части равенства (III,102) также равен нулю. Итак

$$d_{ji}^{\mathrm{T}} K_i^j = \alpha_{ji} y_i^{\mathrm{T}} H_i + \beta_{ji} s_i^{\mathrm{T}} \tag{III, 103}$$

Проводя аналогичные преобразования, можно показать, что

$$c_{ji}^{\mathrm{T}} K_i^j = \gamma_{ji} y_i^{\mathrm{T}} H_i + \delta_{ji} s_i^{\mathrm{T}} \tag{III, 104}$$

Перегруппируем члены в правой части соотношения (II,90)

$$H_{i+1} = H_i + \frac{s_i d_{1i}^{\mathrm{T}} K_i^1}{d_{1i}^{\mathrm{T}} K_i^1 y_i} - H_i \frac{y_i c_{1i}^{\mathrm{T}} K_i^1}{c_{1i}^{\mathrm{T}} K_i^1 y_i} - R y_i \left(\frac{d_{2i}^{\mathrm{T}}}{d_{2i}^{\mathrm{T}} K_i^2 y_i} - \frac{c_{2i}^{\mathrm{T}}}{c_{2i}^{\mathrm{T}} K_i^2 y_i} \right) + R K_i^2 y_i \left(\frac{c_{1i}^{\mathrm{T}} K_i^1}{c_{1i}^{\mathrm{T}} K_i^1 y_i} - \frac{c_{2i}^{\mathrm{T}} K_i^2}{c_{2i}^{\mathrm{T}} K_i^2 y_i} \right) \tag{III, 105}$$

Подставляя в последнюю скобку правой части равенства (III,105) выражения (III,104) для $c_{ji}^{\mathrm{T}} K_i^j$, легко показать, что член в этой скобке будет равен нулю, если константы γ_{ji}, δ_{ji} удовлетворяют соотношению

$$\gamma_{1i} \delta_{2i} = \gamma_{2i} \delta_{1i} \tag{III, 106}$$

В дальнейшем будем исходить из предположения, что константы γ_{1i}, δ_{1i} выбираются произвольно, а константы γ_{2i}, δ_{2i} с учетом выполнения этого равенства. Рассмотрим предпоследнюю скобку в правой части равенства (III,105). Коэффициенты γ_{2i}, δ_{2i} в формуле (III,101) связаны соотношением (III,106), величины же d_{2i}^{T} произвольны. Поэтому мы можем выбрать их таким образом, чтобы член, стоящий в этой скобке, также обратился в нуль. Подставляя теперь в равенство (III,105) выражения (III,103), (III,104) для $d_{1i}^{\mathrm{T}} K_i^1$ и $c_{1i}^{\mathrm{T}} K_i^1$, получим формулу (III,95). Но H_{i+1} из выражения (II,90) удовлетворяет уравнению (III,99), следовательно, и H_{i+1} из формулы (III,95) удовлетворяет уравнению (III,99).

Покажем теперь, что условия (III,98), (III,99) выполняются при $i = 0$. Умножим справа соотношение (III,95) при $i = 0$ на y_0, после несложных преобразований получим $H_1 y_0 = s_0$, что и доказывает правильность соотношения (III,99). Правильность соотношения $s_1^{\mathrm{T}} y_0 = 0$ была доказана (см. [11, с. 63]).

Совокупность конкретных формул для вычисления H_i, которые получаются из выражения (III,95), если давать определенные значения произвольным постоянным α_{ji}, β_{ji}, γ_{ji}, δ_{ji}, для краткости будем называть семейством формул (алгоритмов) Хуанга. Интересно отметить, что в то время, как в формулах (II,90), (II,91) в об-

щем случае требуется запоминать три матрицы (H_i, K_i^1, K_i^2), в формуле Хуанга запоминать надо только одну матрицу H_i. Однако преимущество формул (II,90), (II,91) состоит в том, что они позволяют строить приближения к прямому или обратному гессиану с помощью любой совокупности векторов s_i, y_i (важно только, чтобы они были линейно независимыми). Применение же формулы Хуанга ограничивается случаем, когда направления являются сопряженными. Приведем некоторые известные алгоритмы, которые могут быть получены из формулы (III,95), при различных значениях параметров α_{1i}, β_{1i}, γ_{1i}, δ_{1i}.

Формула Давидона — Флетчера — Пауэлла (*DFP*), $\beta_{1i} = 1$, $\delta_{1i} = 0$, $\alpha_{1i} = 0$, $\gamma_{1i} = 1$, [см. выражение (III,81)].

Формула Маккормика, $\beta_{1i} = 1$, $\alpha_{1i} = 0$, $\delta_{1i} = 1$, $\gamma_{1i} = 0$

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, s_i^{\mathrm{T}}}{s_i^{\mathrm{T}} y_i}$$

Формула Пирсона [31], $\beta_{1i} = 0$, $\alpha_{1i} = 1$, $\delta_{1i} = 0$, $\gamma_{1i} = 1$

$$H_{i+1} = H_i + \frac{(s_i - H_i y_i)\, y_i^{\mathrm{T}} H_i}{y_i^{\mathrm{T}} H_i y_i}$$

Семейство преобразований Бройдена. Если решение H_i матричного уравнения (II,32) искать в классе симметричных матриц (H_i — аппроксимация симметричного обратного гессиана), то преобразованию (III,95) можно придать следующий вид [11, с. 74—75]:

$$H_{i+1} = H_i + \frac{s_i s_i^{\mathrm{T}}}{s_i^{\mathrm{T}} y_i} - \frac{H_i y_i y_i^{\mathrm{T}} H_i}{y_i^{\mathrm{T}} H_i y_i} + \theta_i w_i w_i^{\mathrm{T}} \qquad \text{(III, 107)}$$

где

$$\theta_i = \beta \left(y_i^{\mathrm{T}} H_i y_i\right) s_i^{\mathrm{T}} y_i \qquad w_i = \frac{s_i}{s_i^{\mathrm{T}} y_i} - \frac{H_i y_i}{y_i^{\mathrm{T}} H_i y_i} \qquad \text{(III, 108)}$$

Семейство (III,107)—(III,108) преобразований матрицы H_i с произвольным β рассматривалось Бройденом [67]. Интересно отметить, что преобразования *DFP* и *BFGP*, содержащиеся в уравнениях (III,107), (III,108) соответственно при $\theta_i \equiv 0$ и $\theta_i = y_i^{\mathrm{T}} H_i y_i$ были получены ранее [см. формулы (III,81), (III,84)] с использованием минимизации нормы Фробениуса.

В работах [68; 65] показано, что в определенном смысле формула *BFGS* (III,84) является наилучшей в семействе (III,107) при $\theta_i \geqslant 0$: соответствующие матрицы H_i обладают наибольшей величиной минимального собственного значения и, кроме того, при минимизации квадратичной функции с положительно определенной матрицей G эвклидова норма матрицы E_i, т. е. $|E_i| = Tr\left(E_i^{\mathrm{T}} E_i\right)$, определенной в виде $E_i = BH_iB - I_n$, $B^2 = G$, монотонно уменьшается: $|E_{i+1}| < |E_i|$.

Среди других преобразований следует отметить *SSVM*, входящее в двухпараметрическое семейство Орена [69—71], см. также работу [11, с. 76].

Аппроксимация прямого гессиана. В этом случае матрица B_i должна удовлетворять уравнению (II,29). Аналогично выводу соотношения (III,91) можно показать, что если построены n линейно-независимых направлений $s_0, \ldots, s_{n-1}$, то на n-м шаге будет выполняться равенство

$$B_n = A \qquad \text{(III, 109)}$$

Если направления $s_0, \ldots, s_{n-1}$ сопряженные, то их линейная независимость гарантируется (см. с. 81).

Все что было сказано относительно квазиньютоновских методов 2-го рода для решения систем нелинейных уравнений (см. с. 46), может быть повторено для этого случая, поэтому не будем здесь на этом останавливаться.

Применение квазиньютоновских методов к минимизации произвольных функций

Теория квадратичных методов минимизации, изложенная в начале этой главы, основана на исследовании задачи о минимуме квадратичной функции. Возможность применения этих методов к минимизации произвольных, т. е. неквадратичных функций связана с тем, что при выполнении известных условий неквадратичную функцию в некоторой окрестности точки минимума можно с определенной точностью аппроксимировать квадратичной функцией. Некоторые свойства квадратичных методов минимизации — устойчивость, идентичность генерируемых последовательностей $\{x_i\}$ — установлены, по существу, для неквадратичных минимизируемых функций [67; 72; 11, с. 76—81]. Окончательное решение вопроса о возможности применения квадратичных методов к минимизации неквадратичных функций определяется исследованием сходимости рассматриваемых методов, так как свойство конечности алгоритма (достижение минимума за конечное число итераций) для неквадратичных минимизируемых функций, вообще говоря, не выполняется. Для многих, наиболее часто применяемых квадратичных методов минимизации не только доказано свойство сходимости, но и получены оценки скорости сходимости, которая оказывается сверхлинейной [154, 155]. В то же время метод наискорейшего спуска, например, характеризуется, в общем, более слабой — линейной скоростью сходимости. Практическое подтверждение этих теоретических соображений основывается на результатах решения тестовых задач различными методами и последующей их сравнительной оценке.

В этом разделе на основе тестовых примеров будут приведены некоторые характеристики рассмотренных выше алгоритмов минимизации. Рассмотрим в общих чертах описание некоторых составных частей применяемых квадратичных алгоритмов минимизации.

В главе I были рассмотрены общие принципы построения широкого класса алгоритмов безусловной минимизации и приведена их структурная схема (см. рис. 4). Из этой схемы явствует, что каждый алгоритм минимизации имеет неотъемлемую составную часть — процедуру одномерного спуска. Избранный алгоритм одномерного движения во многом определяет эффективность применяемого алгоритма безусловной минимизации. В качестве процедуры одномерного движения для большинства рассматриваемых здесь алгоритмов минимизации принят процесс вычисления значений минимизируемой функции в последовательно определяемых точках в заданном направлении «спуска»; этот процесс выполняется до того момента, когда не будет найдена первая точка, в которой значение функции меньше, чем в двух соседних точках. Применяемая затем параболическая интерполяция позволяет, в общем случае, улучшить результат, т. е. определить точку с меньшим, чем найденное выше, значением функции. Организованная таким образом процедура линейного поиска дает точное положение минимума ($\alpha_i = \alpha_i^*$) функции в данном направлении p_i при минимизации квадратичной функции. С це-

лью выбора наиболее «подходящей» (с точки зрения сокращения числа последующих вычислений функции) величины начального шага в направлении «спуска» в принятом алгоритме одномерной минимизации предусмотрен анализ процесса линейного поиска, определяющий значение начального шага для каждого следующего направления движения (так называемая «адаптация» начального шага). Для первого направления «спуска» ($p_0 = -g_0$) величина начального шага выбирается с учетом предварительных соображений о характере задачи или же произвольным образом.

В некоторых случаях метод SSVM применялся в сочетании с несколько иным методом одномерного поиска. Величина начального шага вдоль направления p_i зависела от интервала на числовой прямой, в котором оказывалось текущее значение $\|p_i\|$. Если параболическая интерполяция выполнялась в окрестности начальной точки вектора p_i, то учитывалось значение производной минимизируемой функции по направлению p_i в этой точке: (g_i, p_i); интерполяция в этом случае проводилась по двум точкам. Это вело к сокращению числа вычислений функции в процессе поиска. Если же интерполяция выполнялась по трем точкам, то в случае «близости» найденного положения минимума параболы и средней из узловых точек последняя принималась за точку минимума данной функции в направлении p_i. Это также способствовало уменьшению объема вычислений. В приводимых далее табл. 8—16 соответствующие результаты (с двойной точностью) даны в фигурных скобках.

Изложенные методы одномерной минимизации являются по существу методами нулевого порядка, т. е. в точках прямой p_i выполняется расчет лишь значений минимизируемой функции. Если же в каждой точке известны и значения градиента данной функции, то могут быть использованы теоретически более эффективные алгоритмы одномерного поиска, основанные на применении так называемых критериев сходимости. При этом автоматически обеспечивается выполнение условия (III, 163), связанного с устойчивостью алгоритма минимизации. Для обеспечения критерия сходимости в случае его невыполнения на первом шаге обычно используются методы линейной экстраполяции совместно с кубической интерполяцией (см. Приложение 2). По данным решения тестовых задач методы первого порядка требуют в среднем 1,1—1,5 вычислений функции (вместе с градиентом) на направлении по сравнению с 2,5—4 вычислениями при методах нулевого порядка.

Алгоритм одномерного поиска первого порядка входит как составная часть в Switch-метод Флетчера [65], в котором используются преобразования матриц (III, 80) с прямой аппроксимацией гессиана, представленных в факторизованном виде. Результаты тестовых испытаний этого метода даны в табл. 8—17 (строка SW). Следует отметить, что при работе с алгоритмами оптимизации, использующими две матрицы для построения обратного гессиана, например выражения (II, 101), и (II, 102), техника работы с матрицами должна быть аналогична изложенной в главе II. Здесь рассматриваются алгоритмы, использующие одну матрицу преобразования, причем

Т а б л и ц а 8. **Тестовая задача 1 (ROS 2) [156]**

Алгоритм минимизации	K_f	K_p	f^*	$\lvert g^* \rvert$
BFGS	99 (91)	24 (23)	10^{-17} (10^{-12})	10^{-7} (10^{-8})
DFP	106 (108)	26 (25)	10^{-20} (0)	10^{-9} (0)
SSVM	120 (125) {120}	30 (32) {33}	10^{-14} (10^{-12}) {10^{-14}}	10^{-7} (10^{-6}) {10^{-6}}
PRM	164 (133)	38 (30)	10^{-14} (0)	10^{-8} (0)
SW	56 (50)	40 (41)	10^{-19} (0)	10^{-8} (0)

Т а б л и ц а 9. **Тестовая задача 7 [63]**

Алгоритм минимизации	K_f	K_p	f^*	$\lvert g^* \rvert$
BFGS	101 (101)	23 (23)	10^{-18} (10^{-18})	10^{-7} (10^{-7})
DFP	97 (104)	24 (25)	10^{-18} (10^{-16})	10^{-8} (10^{-8})
SSVM	107 (118) {95}	23 (25) {26}	10^{-18} (10^{-11}) {10^{-14}}	10^{-6} (10^{-5}) {10^{-6}}
PRM	149 (186)	32 (38)	10^{-12} (10^{-10})	10^{-5} (10^{-9})
SW	38 (38)	29 (29)	10^{-19} (10^{-21})	10^{-9} (10^{-8})

Т а б л и ц а 10. **Тестовая задача 12 (TEST 10) [159]**

Алгоритм минимизации	K_f	K_p	f^*	$\lvert g^* \rvert$
BFGS	178 (180)	43 (45)	10^{-14} (0)	10^{-6} (0)
DFP	918 (761)	225 (205)	10^{-13} (0)	10^{-6} (0)
SSVM	215 (205) {68}	50 (51) {24}	10^{-13} (10^{-13}) {10^{-13}}	10^{-6} (10^{-6}) {10^{-6}}
PRM	399 (351)	92 (79)	10^{-12} (10^{-12})	10^{-6} (10^{-6})
SW	96 (97)	81 (81)	10^{-17} (0)	10^{-8} (0)

Т а б л и ц а 11. **Тестовая задача 13 (WOOD 4) [160, с. 198]**

Алгоритм минимизации	K_f	K_p	f^*	$\lvert g^* \rvert$
BFGS	156 (161)	42 (42)	10^{-11} (0)	10^{-4} (0)
DFP	184 (199)	49 (51)	10^{-13} (10^{-10})	10^{-5} (10^{-4})
SSVM	120 (120) {86}	26 (26) {22}	10^{-12} (10^{-10}) {10^{-10}}	10^{-4} (10^{-5}) {10^{-4}}
PRM	296 (298)	69 (69)	10^{-10} (10^{-10})	10^{-4} (10^{-4})
SW	99 (107)	77 (80)	10^{-18} (0)	10^{-7} (0)

Таблица 12. **Тестовая задача 14 (WATSON 9) [159]**

Алгоритм минимизации	K_f	K_p	$f*$	$\lvert g*\rvert$
BFGS	159 (123)	33 (25)	$0{,}1399760 \times 10^{-5}$ (10^{-4})	10^{-6} (10^{-3})
DFP	178 (150)	40 (31)	$0{,}1399760 \times 10^{-5}$ (10^{-4})	10^{-8} (10^{-4})
SSVM	301 (171) {207}	62 (38) {64}	$0{,}1399760 \times 10^{-5}$ (10^{-4}) $\{0{,}1399760 \cdot 10^{-5}\}$	10^{-5} (10^{-3}) $\{10^{-6}\}$
PRM	366 (225)	77 (51)	$0{,}1399760 \times 10^{-5}$ (10^{-5})	10^{-6} (10^{-3})
SW	96 (69)	92 (58)	$0{,}1399760 \cdot 10^{-5}$ $(0{,}6656350 \cdot 10^{-5})$	10^{-9} (10^{-4})

Таблица 13. **Тестовая задача 15 (WATSON 6) [159]**

Алгоритм минимизации	K_f	K_p	$f*$	$\lvert g*\rvert$
BFGS	99 (121)	23 (24)	$0{,}2287670 \cdot 10^{-2}$ $(0{,}2287605 \cdot 10^{-2})$	10^{-6} (10^{-3})
DFP	107 (152)	27 (28)	$0{,}2287670 \cdot 10^{-2}$ $(0{,}2287475 \cdot 10^{-2})$	10^{-6} (10^{-3})
SSVM	145 (218) {93}	31 (36) {32}	$0{,}2287670 \cdot 10^{-2}$ $(0{,}2287537 \cdot 10^{-2})$ $\{(0{,}2287670 \cdot 10^{-2})\}$	10^{-6} (10^{-4}) $\{10^{-5}\}$
PRM	192 (229)	40 (37)	$0{,}2287670 \cdot 10^{-2}$ $(0{,}2287670 \cdot 10^{-2})$	10^{-6} (10^{-8})
SW	45 (44)	42 (38)	$0{,}2287670 \cdot 10^{-2}$ $(0{,}2287555 \cdot 10^{-2})$	10^{-8} (10^{-8})

Таблица 14. **Тестовая задача 16 (ROS 8) [161]**

Алгоритм минимизации	K_f	K_p	$f*$	$\lvert g*\rvert$
BFGS	76 (55)	18 (12)	10^{-20} (10^{-26})	10^{-16} (0)
DFP	152 (184)	37 (44)	10^{-18} (10^{-17})	10^{-15} (10^{-13})
SSVM	52 (55) {82}	11 (12) {26}	10^{-15} (10^{-26}) $\{10^{-16}\}$	10^{-13} (0) $\{10^{-13}\}$
PRM	90 (101)	25 (30)	10^{-16} (10^{-15})	10^{-13} (10^{-13})
SW	111 (210)	100 (182)	10^{-24} (10^{-40})	10^{-20} (10^{-34})

Таблица 15. **Тестовая задача 17 (EXP 5) [162]**

Алгоритм минимизации	K_f	K_p	f^*	$\|g^*\|$
BFGS	298 (250)	75 (61)	10^{-15} (10^{-12})	10^{-7} (10^{-6})
DFP	774 (404)	187 (105)	10^{-2} (10^{-12})	10^{-8} (10^{-7})
SSVM	321 (341) {401}	79 (84) {127}	10^{-14} (10^{-14}) {10^{-14}}	10^{-7} (10^{-6}) {10^{-6}}
PRM	599 (1001)	123 (193)	10^{-11} (10^{-12})	10^{-7} (10^{-7})
SW	121 (96)	94 (76)	10^{-2} (10^{-2})	10^{-12} (10^{-3})

Таблица 16. **Тестовая задача 18 (EXP 4) [162]**

Алгоритм минимизации	K_f	K_p	f^*	$\|g^*\|$
BFGS	109 (174)	24 (29)	10^{-20} (10^{-12})	10^{-9} (10^{-7})
DFP	116 (154)	27 (28)	10^{-17} (10^{-12})	10^{-8} (10^{-7})
SSVM	111 (137) {80}	25 (27) {26}	10^{-20} (10^{-14}) {10^{-20}}	10^{-9} (10^{-7}) {10^{-10}}
PRM	159 (197)	35 (37)	10^{-14} (10^{-14})	10^{-8} (10^{-7})
SW	44 (45)	40 (40)	10^{-19} (10^{-13})	10^{-10} (10^{-6})

Таблица 17. **Тестовая задача 19 (WEIBULL) [163]**

Алгоритм минимизации	K_f	K_p	f^*	$\|g^*\|$
BFGS	166 (221)	32 (40)	10^{-17} (10^{-10})	10^{-8} (10^{-6})
DFP	1155 (66;	243 (8)	10^{-19} ($0{,}2134192 \cdot 10^{-1}$)	10^{-10} (10^{-4})
SSVM	253 (121)	50 (14)	10^{-17} ($0{,}2133873 \cdot 10^{-1}$)	10^{-6} (10^{-4})
PRM	184 (465)	26 (57)	10^{-14} (10^{-10})	10^{-8} (10^{-5})
SW	100 (36)	67 (16)	10^{-31} ($0{,}2134119 \cdot 10^{-1}$)	10^{-16} (10^{-4})

Таблица 18. **Тестовая задача 11 при $(x_0)_i \cong 1{,}782$ [152]**

Алгоритм минимизации ($\varepsilon = -15$)	K_f	K_p	f^*
PRM	1914	493	10^{-18}
DFP	414	110	10^{-19}
BFGS	388	104	10^{-19}
SSVM	407	117	10^{-19}
SSVM ($\varepsilon = -7$)	302	91	10^{-12}

Таблица 19. **Результаты решения тестовых задач с увеличением их размерности (n)**

n	Результаты решения тестовых задач 2—10, 20								
	2 [157]	3 [157]	4 [157]	5 [157]	6 [157]	8 [157]	9 [158, с. 143]	10 [158, с. 142]	20 [70]
10	25 18 10^{-21} 10^{-10}	27 20 10^{-23} 10^{-11}	25 19 10^{-23} 10^{-11}	49 44 10^{-18} 10^{-8}	13 6 1 10^{-6}	15 8 1 10^{-6}	43 40 10^{-10} 10^{-8}	13 11 —10 10^{-5}	75 61 10^{-11} 10^{-3}
50	78 60 10^{-23} 10^{-11}	77 60 10^{-22} 10^{-10}	73 62 10^{-21} 10^{-10}	102 82 10^{-16} 10^{-7}	14 5 1 10^{-5}	16 9 1 10^{-6}	51 44 10^{-9} 10^{-7}	57 55 —50 10^{-4}	355 303 10^{-10} 10^{-3}
100	138 115 10^{-20} 10^{-9}	138 115 10^{-20} 10^{-9}	138 116 10^{-20} 10^{-9}	184 157 10^{-15} 10^{-6}	13 6 1 10^{-6}	18 9 1 10^{-4}	68 62 10^{-9} 10^{-7}	110 106 —100 10^{-4}	637 508 3,986614 10^{-3}
150	188 158 10^{-12} 10^{-5}	190 158 10^{-12} 10^{-5}	202 170 10^{-20} 10^{-9}	288 194 10^{-16} 10^{-7}	16 6 1 10^{-5}	16 8 1 10^{-6}	62 57 10^{-8} 10^{-6}	160 157 —150,43 10^{0}	950 748 0 0
200	193 170 10^{-12} 10^{-5}	194 169 10^{-11} 10^{-5}	237 203 10^{-12} 10^{-3}	319 265 10^{-14} 10^{-7}	12 5 1 10^{-6}	17 8 1 10^{-5}	79 73 10^{-9} 10^{-7}	211 205 —201,2 10^{0}	1263 992 10^{-11} 10^{-4}

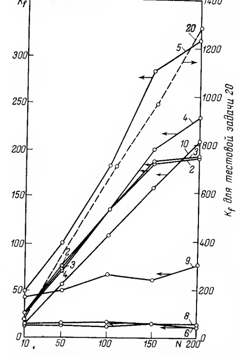


Рис. 13. Кривые роста числа вычислений минимизируемой функции с увеличением размерности тестовых задач: цифры у кривых соответствуют номерам тестовых задач (см. Приложение 1).

в симметричной форме. Среди представленных здесь квадратичных методов минимизации: BFGS (III, 84), DEP (III, 81), PRM (III, 40) и SSVM, использующих алгоритмы одномерного спуска нулевого порядка, наилучшие результаты дал метод BFGS. Вычисления производились как с числами одинарной длины (4 байта), так и при двойной длине слова (8 байтов).

Во всех используемых далее квадратичных алгоритмах минимизации функций многих переменных принят следующий (относительный) критерий окончания решения: порядок отношения норм градиентов в конечной и начальной точках меньше некоторого заданного (отрицательного) числа ε. Если же в процессе решения задачи требуемой точности достигнуть не удается, то окончание работы алгоритма минимизации определяется наперед заданной близостью точек минимума, найденных при линейном «спуске» вдоль каждого из двух последовательных направлений движения.

Среди выбранных тестовых примеров содержатся менее трудные (задачи 1, 7, 13, 16, 18) и более трудные (задачи 12, 14, 15, 17, 19) для минимизации функции. Результаты решения тестовых задач с применением различных (квадратичных) алгоритмов минимизации даны в табл. 8—18; параметр ε в критерии окончания работы алгоритма принят равным —7 всюду, где это специально не указывается; результаты для SSVM, полученные при использовании другого варианта одномерного поиска, обсуждавшегося ранее в этом разделе, приведены в фигурных скобках; результаты с одинарной точностью даны в круглых скобках.

В таблице 19 и на рис. 13 приведены результаты решения тестовых задач 2—6, 8—10, 20 с увеличением размерности n от 10 до 200 (каждый результат представлен четверкой чисел — K_f, K_p, f^*, $|g^*|$.

Решение выполнялось методом SW с одинарной точностью. С ростом n вычислительные затраты K_f увеличиваются, но не быстрее, чем при прямой пропорциональной зависимости.

ГЛАВА IV

Оптимизация химико-технологических процессов при наличии ограничений

Характеристика ограничений в задачах оптимизации

Задача оптимизации химико-технологических процессов по существу сводится к нахождению некоторого компромисса между выбором определенных условий проведения процесса (характер цели) и ограниченностью ресурсов (средства достижения цели). Характер компромисса, принятого при решении конкретной задачи, сказывается на форме критерия оптимизации и в большинстве случаев предполагает наличие явного указания на ограниченность ресурсов определенного вида, например, расходов сырья. Кроме того, при проведении конкретного химико-технологического процесса обычно должны быть выдержаны определенные условия, т. е. ограничения, налагаемые на значения его параметров; эти ограничения связаны с характером принятой технологии и т. п. Ограничения, встречающиеся в задачах оптимизации химико-технологических процессов, можно подразделить на две группы.

К первой группе отнесем те, которые налагаются на зависимые переменные системы. К ним относятся выходные и промежуточные переменные системы, а также фазовые переменные аппаратов, являющихся объектами с распределенными параметрами. Ограничения первой группы появляются в случае:

постановки задачи оптимизации в форме [73, с. 15]: найти минимум затрат при заданной производительности целевого продукта, найти максимум производительности при заданной себестоимости, и т. п.;

наличия ограничений по качеству выходного продукта, т. е. когда требуется, чтобы количества непрореагировавших или побочных продуктов не превышали заданных величин;

наличия ограничений на переменные потока внутри аппаратов (объекты с распределенными параметрами), например, иногда задается ограничение на температуру реакционной смеси в каталитическом реакторе. Можно привести и другие ограничения, относящиеся к первой группе.

Ко второй группе отнесем ограничения, налагаемые на независимые переменные системы: входные переменные и управляющие переменные аппаратов. Их можно подразделить следующим образом:

ограничения на технологические управляющие переменные, обусловленные возможностями (в данном случае мощностью) аппаратуры; так, величина нагрузки на реактор ограничивается мощностью компрессора, температура холодильника реактора — некоторым максимальным значениям, определяемым конструкцией холодильника, температура куба ректификационной колонны — необходимостью предотвратить необратимые изменения разделяемой смеси, и т. д.;

ограничения на конструктивные переменные (например, длина трубки, диаметр реактора);

ограничения, связанные с условиями взрывобезопасности и т. п.

Рассмотрим обе группы переменных с точки зрения учета их в методах оптимизации, при которых схема рассматривается как единое целое. Несмотря на то, что все ограничения можно формально свести к виду (I, 8), в большинстве случаев они совершенно различны по своим свойствам. Остановимся на этом подробнее. Начнем с методов первой группы. Они обладают следующими свойствами:

явный вид функции φ_i [см. выражение (I, 8)] нам неизвестен, мы можем только найти ее численное значение, зная управления в блоках и рассчитав схему;

ограничения первой группы являются, по существу, принципиально нелинейными, поскольку модели блоков как правило нелинейны;

ограничения на выходные переменные k-го блока эквивалентны наличию ограничений на управления всех блоков, входящих в зону воздействия [3, с. 18] на k-тый блок; так, наличие ограничений на выходные переменные последнего блока простой последовательности блоков (см. рис. 22) эквивалентно наличию ограничений на управления всех блоков данной последовательности.

Таким образом, ограничения первой группы носят глобальный характер, связывая воедино переменные многих блоков.

Рассмотрим теперь свойства ограничений второй группы:

они носят локальный, поблочный характер поскольку налагаются на совокупности управлений, относящихся к отдельным блокам схемы;

их явный вид известен;

во многих случаях они линейны и имеют вид либо выражения (IV, 98), либо еще более простой [см. неравенство (I, 9)].

Присутствие ограничений первой группы существенно усложняет задачу оптимизации. В этой главе будут рассмотрены методы последовательной безусловной минимизации и методы с непосредственным учетом ограничений. Последние представлены методом обобщенного приведенного градиента (МОПГ) и рядом методов с линейными ограничениями.

Методы последовательной безусловной минимизации

Пусть в n-мерном пространстве переменной x фиксировано множество Ω, на котором определены функция $f(x)$ и отображение $h(x)$ в пространство E^m. Задача заключается в определении минимума функции

$$\min f(x) \qquad \text{(IV, 1)}$$

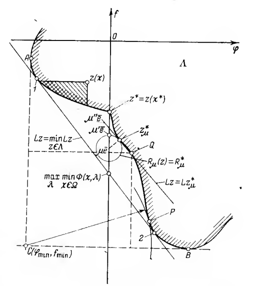


Рис. 14. Геометрическая интерпретация задачи с ограничениями.

при условии

$$h(x) \in \Gamma \leqslant E^m \qquad \text{(IV, 2)}$$

В большинстве практических задач $h(x)$ в условии (IV,2) записывается в виде

$$\varphi_i(x) = 0,$$

$$i = 1, \ldots, p \qquad \text{(IV, 3)}$$

$$\psi_j(x) \leqslant 0, \; j = 1, \ldots, q$$

$$m = p + q \qquad \text{(IV, 4)}$$

Ограничения типа неравенства (IV, 4) могут быть формально представлены как ограничения типа равенства (IV, 3), для этого вводят соответствующие функции

$$\varphi_{j+p}(x) = \psi_j(x)\, H(\psi_j^*(x)) = 0 \qquad j = 1, \ldots, q \qquad \text{(IV, 5)}$$

где $H(\xi)$ — функция Хевисайда; при $\xi > 0$ $H(\xi) = 1$, при $\xi \leqslant 0$ $H(\xi) = 0$. В дальнейшем мы будем иметь дело только с этой, так называемой, канонической формой [см. выражения (IV, 1), (IV, 3), (IV, 5)] задачи на условный экстремум.

Пусть решение x^* исходной задачи существует. Введем пространство $Z\ (=E^{m+1})$, содержащее значения функций $(f(x), \varphi_1(x), \ldots, \varphi_p(x), \varphi_{p+1}(x), \ldots, \varphi_m(x))$, которое будем рассматривать как прямое произведение пространства E^1 значений функции $f(x)$ и пространства $\Psi\ (=E^{p+q})$ значений функций φ_i, $i = 1, \ldots, m$, определяющих ограничения (IV, 3), (IV, 5). Пространство Z использовалось для решения различных экстремальных задач [74, 75]. Элементы $z \in Z$ имеют вид: $z = (f, \varphi)$, где f — число, а $\varphi \in \Psi$. Обозначим через $0_Z = (0, 0_\Psi)$ нулевой элемент пространства Z, а через $\bar{e} = (1, 0_\Psi)$ — единичный вектор в направлении оси Of. Введем множество

$$\Lambda = \{z \mid z \in Z,\ z = z(x),\ x \in \Omega\} \qquad \text{(IV, 6)}$$

где $z(x) = [f(x), \varphi(x)]$. Элемент $z(x)$ является образом элемента x в пространстве Z, а множество Λ — образом множества Ω (рис. 14). Заметим, что обратное отображение, переводящее множество Λ в Ω в общем случае не является однозначным.

Выберем в пространстве переменной x элемент $x \in \Omega$, удовлетворяющий ограничениям (IV, 3), (IV, 5), т. е. такой, что $\varphi(x) = 0_\Psi$. Образ этого элемента $z(x) = [f(x), 0_\Psi]$ в пространстве Z является точкой оси Of. Отсюда ясно, что образом множества точек $x \in \Omega$, которые удовлетворяют ограничениям (IV, 3), (IV, 5) задачи, является пересечение множества Λ с осью Of, а образ $z^* = z(x^*) =$

$= [f(x^*), \varphi(x^*)]$ оптимальной точки x^*, т. е. решения задачи (IV, 1), (IV, 3), (IV, 4), является «нижней» точкой этого пересечения, т. е. точкой (f, φ) множества $\Lambda \cap Of$, имеющей наименьшее значение компоненты f. Очевидно, что точка $z(x^*)$ лежит на границе множества Λ. В противном случае (если бы она являлась внутренней точкой Λ) можно было бы указать элемент множеста Λ, расположенный на оси Of «ниже» точки $z(x^*)$, что противоречит предположению об оптимальности x^*.

Установленный таким образом факт, что точка $z(x^*)$ является крайней точкой пересечения оси Of с множеством Λ, позволяет сформулировать для ее определения некоторые задачи оптимизации в пространстве Z значений функций критерия и ограничений. Методы решения этих задач составляют суть так называемой, последовательной безусловной минимизации.

Метод множителей Лагранжа

Рассмотрим задачу минимизации (IV, 1) при ограничениях типа (IV, 3), полагая $\Omega = E^n$. Пусть множество Λ является замкнутым и выпуклым. Введем функцию Лагранжа

$$\Phi(x, \lambda) = f(x) - \lambda^{\mathrm{T}}\varphi(x) \qquad \text{(IV, 7)}$$

где λ — совокупность множителей Лагранжа. В пространстве Z правая часть выражения (IV, 7) определяет линейный функционал $L : Lz = f - \lambda^{\mathrm{T}}\varphi$, причем $L\bar{e} = 1 > 0$. Множество

$$Lz = c \qquad \text{(IV, 8)}$$

в пространстве Z — представляет собой плоскость, которая пересекает ось Of в точке с ординатой, равной величине c (рис. 15). Если в семействе определяемых функционалом L параллельных плоскостей, пересекающих множество Λ, выбрать плоскость, обладающую самой «низкой» точкой пересечения с осью Of, т. е. если значение c в равенстве (IV, 8) принять равным

$$c' = \min_{z \in \Lambda} Lz \qquad \text{(IV, 9)}$$

то равенство (IV, 8) определяет, так называемую опорную плоскость к множеству Λ (напомним, что плоскость называется опорной к Λ в некоторой его точке, если множество Λ целиком лежит по одну сторону этой плоскости). Находим

$$c' = \min_{z \in L} Lz \leqslant Lz^* = f(x^*) \qquad \text{(IV, 10)}$$

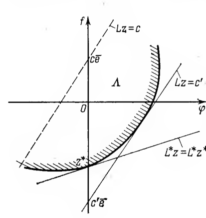


Рис. 15. Геометрическая интерпретация метода множителей Лагранжа.

т. е. точка пересечения любой опорной плоскости, описываемой уравнением (IV, 8) (в котором $c = c'$), с осью Of расположена во всяком случае не «выше» точки z^*. Другими словами, величины c' являются оценками $f(x^*)$ снизу, причем их можно использовать для нахождения точки $z^* = z(x^*)$, а следовательно, и решения x^* исходной задачи.

Если $-\infty < c < f(x^*)$ $(c\bar{e} \notin \Lambda)$, то в силу теоремы об отделимости [76, с. 42] существует (ненулевой) линейный функционал $\tilde{L}$, такой что

$$\tilde{L}(c\bar{e}) < \min_{z \in \Lambda} \tilde{L}z \qquad \text{(IV, 11)}$$

При этом $\tilde{L}\bar{e} = \lambda_0$, $\lambda_0 \neq 0$, так как в противном случае $\tilde{L}(c\bar{e}) = c \cdot \tilde{L}\bar{e} = 0$ и $\tilde{L}z^* = \tilde{L}[f(x^*)\bar{e}] = f(x^*)\tilde{L}\bar{e} = 0$, т. е. точки $(c, 0_\Psi)$ и $z^* \in \Lambda$ не разделяются функционалом $\tilde{L}$, что противоречит неравенству (IV, 11). Не ограничивая общности, можно принять, что в неравенстве (IV, 11) $\lambda_0 = 1$, т. е. что функционал $\tilde{L}$ имеет вид $\tilde{L}z = f - \tilde{\lambda}^\tau \varphi$. Соотношение (IV, 11) означает, что точка $(\tilde{c}, 0_\Psi)$ пересечения опорной к Λ плоскости, определяемой функционалом $\tilde{L}$, с осью Of лежит «выше» точки $(c, 0_\Psi)$:

$$\tilde{L}(c\bar{e}) = c < \tilde{c} = \min_{z \in \Lambda} \tilde{L}z \leqslant f(x^*) \qquad \text{(IV, 12)}$$

В то же время, так как z^* не является внутренней точкой Λ, существует опорная к Λ плоскость [76, с. 41], проходящая через z^*, т. е. существует ненулевой линейный функционал L^* над Z, такой, что

$$L^*z^* = \min_{z \in \Lambda} L^*z \qquad \text{(IV, 13)}$$

Если предполагать, что x^* является регулярной точкой отображения $\varphi: E^n \to E^p$, определяемого функциями левой части равенства (IV, 3), т. е. линейный оператор $\partial\varphi(x^*)/\partial x$ отображает E^n на все пространство E^p или, что то же, векторы

$$\frac{\partial \varphi_1(x^*)}{\partial x}, \ldots, \frac{\partial \varphi_p(x^*)}{\partial x}$$

линейно независимы, то $L^*\bar{e} \neq 0$. Действительно, если $L^*\bar{e} = \lambda_0 = 0$, то необходимое условие минимума функционала L^*z на множестве Λ [для левой части выражения (IV, 13)] или, что то же, минимума $L^*z(x)$ $[L^*z(x) = \lambda_0 f(x) - \lambda^\tau \varphi(x)]$ для $x \in E^n$:

$$\lambda_0 \frac{\partial f(x^*)}{\partial x} - \left(\frac{\partial \varphi(x^*)}{\partial x}\right)^\tau \lambda = 0 \qquad \text{(IV, 14)}$$

вместе с условием регулярности дает $\lambda = 0$, т. е. функционал L^* оказывается нулевым. Полученное противоречие показывает, что в выражении (IV, 13), не ограничивая общности, можно принять

$\lambda_0 = 1$ и, следовательно, функционал L^*z взять в виде: $L^*z = f - \lambda^{\tau}\varphi$. Отсюда следует, что

$$L^*z^* = \max_L c' = \max_L \min_{z \in \Lambda} Lz = \max_{\lambda} \min_{x \in \Omega} \Phi(x, \lambda) \qquad (\Omega = E^n)$$

т. е.

$$f(x^*) = \Phi(x^*, \lambda^*) = \max_{\lambda} \min_{x \in \Omega} \Phi(x, \lambda) \tag{IV, 15}$$

Заметим, что поскольку в предшествующем изложении речь шла о построении опорных плоскостей, минимум

$$\min_{x \in \Omega} \Phi(x, \lambda)$$

в выражении (IV, 15) следует понимать как глобальный минимум функции $\Phi(x, \lambda)$ на множестве Ω при фиксированном векторе $\lambda \in E^p$. Из рис. 15 ясно, что операция взятия максимума по λ в выражении (IV, 15) также представляет глобальный максимум по λ. Это легко показать и аналитически. Действительно, пусть $x(\lambda) \in \Omega$ является точкой глобального минимума $\Phi(x, \lambda)$, т. е.

$$\Phi[x(\lambda), \lambda] = \min_{x \in \Omega} \Phi(x, \lambda)$$

Тогда для любого λ

$$\Phi[x(\lambda), \lambda] \leqslant \Phi(x^*, \lambda) \tag{IV, 16}$$

Так как $\varphi(x^*) = 0$, из формулы (IV,7), находим

$$\Phi(x^*, \lambda) = \Phi(x^*, \lambda^*) \tag{IV, 17}$$

Следовательно

$$\Phi[x(\lambda), \lambda] \leqslant \Phi(x^*, \lambda^*) = \Phi[x(\lambda^*), \lambda^*] \tag{IV, 18}$$

т. е. λ^* является точкой глобального максимума функции $\Phi(x, (\lambda), \lambda)$, что и выражает соотношение (IV, 15), которое можно записать в виде

$$\Phi(x^*, \lambda^*) = \max_{\lambda} \Phi[x(\lambda), \lambda] \tag{IV, 19}$$

В соотношении (IV, 15), как в частном случае, находит свое выражение общий принцип построения рассматриваемых в данном параграфе методов решения задач с ограничениями. Он предусматривает организацию двухуровневого вычислительного процесса, верхний уровень которого представляет собой итерационную процедуру относительно некоторой совокупности искусственно введенных параметров: в выражении (IV, 15) — это поиск максимума функции $\Phi[x(\lambda), \lambda)]$ относительно множителей Лагранжа λ, а нижний — итерационную процедуру в пространстве независимых переменных x исходной задачи с ограничениями. Эта процедура выполняется при фиксированных значениях «искусственных» параметров верхнего уровня: в выражении (IV, 15) — это минимизация по переменным $x \in \Omega$ функции Лагранжа (IV, 17) при фиксированных λ. В основном нам придется иметь дело со специальным типом двухуровневого процесса, аналогичным описываемому выражением (IV, 15), в котором процедура верхнего уровня представляет собой поиск максимума некоторой функции по известным параметрам, а процедура нижнего уровня — поиск ее минимума по переменным x.

Рассмотрим более подробно верхний уровень процесса (IV, 15), т. е. выражение (IV, 19). В дополнение к ситуации, изображенной на рис. 15, в соответствии с которой функция (IV, 17) имеет минимум на множестве $\Omega = E^n$ в некоторой точке $x(\lambda)$, будем предполагать выполненным достаточное условие минимума функции $\Phi(x, \lambda)$ в точке $x(\lambda)$:

$$z^T V[x(\lambda), \lambda] z > 0 \qquad \text{(IV, 20)}$$

где $z \in E^n$, $z \neq 0$, а

$$V[x(\lambda), \lambda] = \left(\frac{\partial^2 \Phi[x(\lambda), \lambda]}{\partial x_i \partial x_j} \right)$$

есть матрица вторых производных функции $\Phi(x, \lambda)$ в точке $x(\lambda)$. [Для выполнения условия (IV, 20) достаточно потребовать его выполнения в точке $x^* = x(\lambda^*)$ и непрерывности вторых производных $\partial^2\Phi/\partial x_i \partial x_j$.]

Вычислим первые и вторые производные функции $\overline{\Phi}(\lambda) = \Phi[x(\lambda), \lambda]$. Находим

$$\frac{\partial \overline{\Phi}}{\partial \lambda} = \left(\frac{\partial x}{\partial \lambda} \right)^T \frac{\partial \Phi}{\partial x} + \frac{\partial \Phi}{\partial \lambda} \qquad \text{(IV, 21)}$$

Но в точке $x(\lambda)$ минимума $\Phi(x, \lambda)$ частная производная $\partial\Phi/\partial x$ равна нулю. Учитывая равенство (IV, 7), получаем $\partial\Phi/\partial\lambda = -\varphi(x)$ и, следовательно

$$\frac{\partial \overline{\Phi}}{\partial \lambda} = -\varphi[x(\lambda)] \qquad \text{(IV, 22)}$$

Последующее дифференцирование этого соотношения по λ дает

$$\nabla_\lambda^2 \overline{\Phi} = -\frac{\partial \varphi}{\partial x} \cdot \frac{\partial x}{\partial \lambda} \qquad \text{(IV, 23)}$$

Производные $\partial x/\partial\lambda$ можно вычислить с помощью тождества $\frac{\partial \Phi[x(\lambda), \lambda]}{\partial x} \equiv 0$. Дифференцируя его, находим:

$$\frac{\partial}{\partial \lambda} \left(\frac{\partial \Phi[x(\lambda), \lambda]}{\partial x} \right) = \nabla_x^2 \Phi \frac{\partial x}{\partial \lambda} + \left[\frac{\partial}{\partial x} \left(\frac{\partial \Phi}{\partial \lambda} \right) \right]^T =$$

$$= \nabla_x^2 \Phi \frac{\partial x}{\partial \lambda} - \left(\frac{\partial \varphi}{\partial x} \right)^T = 0 \qquad \text{(IV, 24)}$$

Отсюда получаем

$$\frac{\partial x}{\partial \lambda} = \left(\nabla_x^2 \Phi \right)^{-1} \left(\frac{\partial \varphi}{\partial x} \right)^T \qquad \text{(IV, 25)}$$

(Обратная матрица существует, так как согласно неравенству (IV, 20) $\Delta_x^2\Phi = V$ положительно определена.) Совместное рассмотрение выражений (IV, 25) и (IV, 23) дает

$$\nabla_\lambda^2 \overline{\Phi} = -\frac{\partial \varphi[x(\lambda)]}{\partial x} \left\{ \nabla_x^2 \Phi[x(\lambda), \lambda] \right\}^{-1} \left(\frac{\partial \varphi[x(\lambda)]}{\partial x} \right)^T = -UV^{-1}U^T \qquad \text{(IV, 26)}$$

где

$$U = \frac{\partial \varphi[x(\lambda)]}{\partial x}.$$

Соотношение (IV, 22) показывает, что в точке $\lambda = \lambda^*$ функция $\Phi [x(\lambda), \lambda]$ действительно имеет экстремум:

$$\frac{\partial \overline{\Phi}(\lambda^*)}{\partial \lambda} = -\varphi [x(\lambda^*)] = -\varphi(x^*) = 0$$

согласно соотношению (IV, 19) — это максимум, а из соотношения (IV, 26) следует, что матрица вторых производных функций $\Phi [x(\lambda), \lambda]$ отрицательно определена, так как $\{\nabla_x^2 \Phi [x(\lambda), \lambda]\}^{-1}$ — положительно определена и векторы градиентов $\partial\varphi_i/\partial x$ при $i = 1, \ldots, p$ линейно независимы (точка $x(\lambda)$ — регулярная). Это последнее обстоятельство позволяет применить метод Ньютона на верхнем уровне, соответствующем операции поиска максимума функции $\Phi [x(\lambda), \lambda]$ в выражении (IV, 15):

$$\lambda_{i+1} = \lambda_i - [\nabla_\lambda^2 \overline{\Phi}(\lambda_i)]^{-1} \frac{\partial \overline{\Phi}(\lambda_i)}{\partial \lambda} = \lambda_i - (UV^{-1}U^{\text{т}})^{-1} \varphi [x(\lambda_i)] \qquad (\text{IV}, 27)$$

который, как известно, обладает квадратичной скоростью сходимости в окрестности λ^*. Строго говоря, для применения выражения (IV, 27), необходим расчет элементов матрицы $V [x(\lambda), \lambda]$, т. е. вторых производных (по x) функции $\Phi(x, \lambda)$ в точке $x(\lambda)$, обычно требующий значительного объема вычислений. Однако если на нижнем уровне, соответствующем операции поиска минимума $\Phi(x, \lambda)$ по переменным x в выражении (IV, 15), минимизация выполняется с применением одного из алгоритмов квазиньютоновского типа с аппроксимацией обратного гессиана, то в точке минимума $x(\lambda)$ оказывается известной достаточно хорошая оценка для матрицы V^{-1}, которая и может быть использована в выражении (IV, 27).

Таким образом, с учетом принятых в этом разделе допущений, метод множителей Лагранжа может быть представлен в форме алгоритма, состоящего из следующих шагов.

Алгоритм III.

1. Даны $x_0 \in E^n$, $\lambda_0 \in E^p$, $\varepsilon > 0$. Положить $k = 0$.
2. Найти точку x_{k+1} минимума функции (IV, 7) $\Phi(x, \lambda_k)$.
3. Вычислить «новые» значения λ_{k+1} множителей Лагранжа по формуле (IV, 27).
4. Если

$$\left| \frac{\partial \Phi(x_{k+1}, \lambda_{k+1})}{\partial x} \right| \leqslant \varepsilon$$

то остановиться (x_{k+1} — оптимальная точка); в противном случае положить $k = k + 1$ и перейти к шагу 2.

Сделаем несколько замечаний к приведенному алгоритму.

Шаг 2 представляет собой вычислительную процедуру нижнего уровня — алгоритм безусловной минимизации функции $\Phi(x, \lambda_k)$. Соответствующие методы (первого порядка), обладающие свойством квадратичного завершения (в том числе квазиньютоновские), подробно описаны в главе III. В качестве начального приближения к x_{k+1} здесь и далее принимается точка x_k. Условие окончания работы

алгоритмов этого типа основано на оценке градиента минимизируемой функции, которая в данном случае имеет вид

$$\left| \frac{\partial \Phi(x_{k+1}, \lambda_k)}{\partial x} \right| < \varepsilon_1$$

с некоторым $\varepsilon_1 > 0$.

На шаге 4 выполняется анализ условия остановки алгоритма. Его выполнение означает, что в начальной точке x_{k+1}, предназначенной для последующего выполнения шага 2 с «новыми» значениями λ, уже выполнено (при $\varepsilon = \varepsilon_1$) условие окончания шага 2, т. е. условие окончания работы алгоритма безусловной минимизации, и основной алгоритм должен быть остановлен. Нетрудно установить связь между условием шага 4 и более «естественными» условиями остановки метода множителей: $|\lambda_{k+1} - \lambda_k| < \varepsilon_2$ с некоторым $\varepsilon_2 > 0$ или

$$\left| \frac{\partial \Phi[x(\lambda_k), \lambda_k]}{\partial \lambda_k} \right| < \varepsilon_3$$

с некоторым $\varepsilon_3 > 0$. В самом деле, выражения (IV, 7), (IV, 27) и (IV, 22) дают:

$$\frac{\partial \Phi(x_{k+1}, \lambda_{k+1})}{\partial x} = \frac{\partial f(x_{k+1})}{\partial x} - \left(\frac{\partial \varphi(x_{k+1})}{\partial x} \right)^{\mathrm{T}} \lambda_{k+1} = \frac{\partial \Phi(x_{k+1}, \lambda_k)}{\partial x} -$$

$$- \left(\frac{\partial \varphi[x(\lambda_k)]}{\partial x} \right)^{\mathrm{T}} (\lambda_{k+1} - \lambda_k) = \frac{\partial \Phi(x_{k+1}, \lambda_k)}{\partial x} + U^{\mathrm{T}} (UV^{-1}U^{\mathrm{T}})^{-1} \varphi[x(\lambda_k)] =$$

$$= \frac{\partial \Phi[x(\lambda_k), \lambda_k]}{\partial x} - U^{\mathrm{T}} (UV^{-1}U^{\mathrm{T}})^{-1} \frac{\partial \Phi[x(\lambda_k), \lambda_k]}{\partial \lambda} \qquad \text{(IV, 28)}$$

Это позволяет перейти от одной формы условия остановки к другой ввиду линейной независимости векторов $\partial \varphi_i / \partial x$. Обычно условие шага 4 усиливается требованием достаточной близости точек x_{k+1} и x_k и (или) другими условиями.

В заключение сделаем два замечания. Первое касается принятого здесь предположения о выпуклости множества Λ. Оно является существенным. Действительно, при наличии в окрестности решения так называемого «зазора» между Λ и его выпуклой оболочкой (см. рис. 14), приведенный выше алгоритм, определяющий $\max\limits_{\lambda} \min\limits_{x} \Phi(x, \lambda)$, даст в качестве решения точки x, которым в пространстве Z соответствуют точки *1* или *2* $[\varphi(x) \neq 0]$, в то время как действительному решению соответствует точка z^*. Снять требование выпуклости множества Λ удается при использовании, так называемой «модифицированной» функции Лагранжа, к рассмотрению которой мы обратимся в конце этого раздела. Второе замечание связано с выполнением некоторой процедуры минимизации функции $\Phi(x, \lambda_k)$ на шаге 2 алгоритма метода множителей. Применяемые для этой цели алгоритмы носят, как правило, локальный характер. В этой ситуации метод множителей может привести к локальному решению $x^* \in E^n$ исходной задачи, либо вовсе не дать решения.

Метод „уровней“

Рассмотрим общую задачу минимизации (IV, 1), (IV, 3), (IV, 5), предполагая множество Λ лишь замкнутым (выпуклость Λ необязательна).

Как и в предыдущем случае, изложение будем вести с привлечением двойственных переменных — функционалов над Z.

Пусть $-\infty < \mu < \mu^* = f(x^*)$. Возьмем однопараметрическое (зависящее от параметра μ) семейство определенных на Λ функций $R_\mu(z)$

$$R_\mu(z) = |z - \mu \bar{e}|, \tag{IV, 29}$$

определяющих расстояние от точки $z \in \Lambda$ до точки $\mu\bar{e} \notin \Lambda$.

Обозначим через R^*_μ ($R^*_\mu > 0$) расстояние от точки $\mu\bar{e}$ до множества Λ:

$$R^*_\mu = \inf_{z \in \Lambda} |z - \mu\bar{e}|, \tag{IV, 30}$$

Поскольку Λ — замкнуто, существует $z^*_\mu \in \Lambda$ [149, с. 22], для которого достигается нижняя грань в (IV, 30), т. е. функция (IV, 29) достигает в точке $z^*_\mu = (f^*_\mu, \varphi^*_\mu)$ своего наименьшего значения:

$$R^*_\mu = R_\mu(z^*_\mu) = |z^*_\mu - \mu\bar{e}| = \min_{z \in \Lambda} |z - \mu\bar{e}| > 0 \tag{IV, 31}$$

Теперь можно улучшить нижнюю оценку μ величины $f(x^*)$, положив

$$\mu'(R^*_\mu) = \mu + R^*_\mu \tag{IV, 32}$$

Действительно, в силу равенства (IV, 30) можно записать:

$$\mu < \mu' = \mu + R^*_\mu \leqslant \mu + (\mu^* - \mu) = \mu^* \tag{IV, 33}$$

Таким образом, соответствующий двухуровневый вычислительный процесс для метода «уровней» (см. предыдущий метод) включает определение точной верхней грани, т. е. $\sup\limits_{\mu \leqslant \mu^*} \mu'(R^*_\mu)$ при $\mu'(R^*_\mu)$ из формулы (IV, 32), на в е р х н е м уровне и минимизацию функции (IV, 29) — определение R^*_μ — на н и ж н е м уровне.

Если Λ выпукло, то по заданной величине μ можно получить лучшую, чем μ' оценку

$$\mu'' = \mu + \frac{(R^*_\mu)^2}{f^*_\mu - \mu} \geqslant \mu', \tag{IV, 34}$$

которая определяется как асбцисса точки пересечения с Of опорной к Λ плоскости, проходящей через точку z^*_μ.

В некоторых случаях эту оценку можно получить и при более слабых предположениях, что показано на рис. 14; точка $\mu''\bar{e}$ [μ'' соответствует формуле (IV, 34)] лежит на пересечении оси Of с плоскостью, определяемой некоторым линейным функционалом L, касательной к Λ в точке z^*_μ. Точка $\mu'\bar{e}$ [μ' соответствует формуле (IV, 32)] лежит на пересечении Of со сферой (радиус R^*_μ, центр в точке $\mu\bar{e}$), соприкасающейся с границей множества Λ в точке z^*_μ. Видно, что точка $\mu''\bar{e}$ лежит «выше» точки $\mu\bar{e}$ и одновременно «ниже» точки $z(x^*)$.

Обычно вместо функции (IV, 29) рассматривают ее квадрат

$$F_\mu(x) = R_\mu^2[z(x)] = [f(x) - \mu]^2 + \varphi^2(x) \qquad \text{(IV, 35)}$$

и вводят положительные коэффициенты при члене φ^2:

$$F_{\mu,\sigma}(x) = [f(x) - \mu]^2 + \sum_{i=1}^{p} \alpha_i \varphi_i^2(x) + \sum_{j=1}^{q} \beta_j \psi_j^2(x) H[\psi_j(x)] \qquad \text{(IV, 36)}$$

$$\sigma = (\alpha, \beta) \qquad \alpha_i > 0 \qquad i = 1, \ldots, p \qquad \beta_j > 0 \qquad j = 1, \ldots, q$$

Подобная функция рассматривалась в работе [77]; там же предложен один из алгоритмов определения последовательных значений параметра μ.

Если $\sigma' = K\sigma$, $K > 1$, то находим

$$F^*_{\mu,\sigma'} = \min_{x \in \Omega} F_{\mu,\sigma'}(x) = F_{\mu,\sigma'}(x^*_{\mu,\sigma'}) \geqslant$$

$$\geqslant F_{\mu,\sigma}(x^*_{\mu,\sigma'}) \geqslant \min_{x \in \Omega} F_{\mu,\sigma}(x) = F^*_{\mu,\sigma} \qquad \text{(IV, 37)}$$

Отсюда следует, в частности, что использование $\alpha_i > 1$ при $i = 1, \ldots, p$; $\beta_j > 1$; $j = 1, \ldots, q$ в выражении (IV, 36) позволяет получить лучшую оценку μ' в формуле (IV, 32). Кроме того, при получении очередного значения μ' нижней оценки $f(x^*)$ целесообразно одновременно увеличивать коэффициенты α_i, β_j. Это ясно и геометрически из графика (рис. 16), на котором показана сфера (радиус F^*_μ, центр в точке $\mu\bar{e}$), соприкасающаяся с границей множества Λ и представляющая собой поверхность уровня функции (IV, 36) при $\alpha_i = \beta_j = 1$, обладающую минимальным в силу равенства (IV, 31) значением F^*_μ — уровня на множестве Λ. Точка $\mu'\bar{e}$ лежит на пересечении сферы с осью Of. Пунктиром показан соприкасающийся с границей множества Λ эллипсоид, который представляет собой поверхность уровня функции (IV, 36) при $\sigma > 1$, обладающую минимальным значением уровня $F^*_{\mu\,\sigma}$ на множестве Λ. Эллипсоид вытянут вдоль оси Of. «Верхняя» точка его пересечения с осью расположена «выше» точки $\mu'\bar{e}$, так как в противном случае эллипсоид целиком располагался бы внутри соприкасающейся сферы (за исключением, быть может, своих крайних точек на оси Of), вследствие чего не обладал бы общей точкой с множеством Λ, т. е. не являлся бы «соприкасающимся» эллипсоидом.

Приведенные в этом пункте результаты можно представить в виде алгоритма, состоящего из следующих шагов.

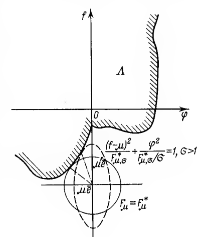


Рис. 16. Геометрическая интерпретация метода уровней.

Алгоритм IV.

1. Даны $x_0 \in E^n$, $\mu_0 < \mu^*$, $\sigma_0 = (\alpha_0, \beta_0)$ [см. формулу (IV, 36)], $\varepsilon > 0$. Положить $k = 0$.

2. Найти точку x_{k+1} минимума функции $F_{\mu_k, \sigma_k}(x)$ [см. формулу (IV, 36)].

3. Вычислить «следующее» значение параметра μ [см. выражения (IV, 32), (IV, 35)]:

$$\mu_{k+1} = \mu_k + \sqrt{F_{\mu_k, \sigma_k}(x_{k+1})}$$

и штрафные коэффициенты $\sigma_{k+1} = K\sigma_k$, $\sigma_{k+1} \leqslant \sigma_{\max}$ (обычно принимают $K = 10$).

4. Если $\mu_{k+1} - \mu_k < \varepsilon$, то остановиться ($x_{k+1}$ — оптимальная точка); в противном случае положить $k = k + 1$ и перейти к шагу 2.

На шаге 1 приведенного алгоритма начальные значения параметра μ и штрафных коэффициентов $\sigma = (\alpha, \beta)$ обычно выбираются на основе имеющейся предварительной информации о данной задаче. Часто оказывается известной нижняя оценка значений критерия оптимизации, позволяющая правильно выбрать μ_0, не занижая его слишком сильно. Характер ограничений, их влияние на поведение модели и требуемая точность выполнения обычно позволяют выбрать подходящие значения для штрафных коэффициентов σ_0. Формула для изменения параметра μ на шаге 3 была предложена в работе [74]. При этом минимум функции $F_{\mu_k, \sigma_k}(x)$ должен быть найден с высокой точностью, иначе для некоторого значения k может оказаться, что $\mu_{k+1} > \mu^*$, т. е. μ_{k+1} уже не будет являться нижней оценкой $f(x^*)$, и последующее многократное применение формулы шага 3 при продолжении работы алгоритма даст либо возрастающую последовательность значений μ, либо приведет к окончанию работы при значении критерия $> f(x^*)$. Версия этого алгоритма, предусматривающая нахождение верхних оценок для μ, в которой наряду с формулой (IV, 32) используют и оценку (IV, 34), предложена в работе [78], см. также [11, с. 157].

Используемое на шаге 4 условие остановки алгоритма по существу является оценкой величины $\sqrt{F_{\mu_k, \sigma_k}(x_{k+1})}$, в которую входят [см. функцию (IV, 36)] отличающиеся одна от другой по своему характеру функции: критерий оптимизации и разного рода ограничения. Применение одной лишь этой оценки обычно сопровождается повышенными требованиями к точности расчета, так как выбор значения $\varepsilon > 0$ зависит от одного или нескольких жестко заданных ограничений или (и) требуемой точности определения оптимального значения критерия. Поэтому условие остановки алгоритма целесообразно применять в следующей форме:

$$\left| \frac{f(x_{k+1-f})(x_k)}{f(x_k)} \right| < \varepsilon_1 \quad \text{или} \quad |f(x_{k+1}) - f(x_k)| < \varepsilon_1 \qquad \varepsilon_1 > 0$$

$$|\varphi_i(x_{k+1})| < \varepsilon_{2_i} \qquad \varepsilon_{2_i} > 0 \qquad i = 1, \ldots, p \qquad \text{(IV, 38)}$$

$$\varphi_{j+p}(x_{k+1}) < \varepsilon_{3_j} \qquad \varepsilon_{3_j} > 0, \qquad j = 1, \ldots, q$$

$$|x_{k+1} - x_k| < \varepsilon_4 \qquad \varepsilon_4 > 0$$

Метод „штрафов“

Если метод множителей Лагранжа можно трактовать геометрически как метод касательных (к границе Λ) плоскостей, а метод «уровней» — как метод соприкасающихся сфер (или эллипсоидов), то традиционный метод штрафных функций можно назвать методом соприкасающихся параболоидов. В этом методе решения задачи (IV, 1), (IV, 3), (IV, 5) рассматривается следующее однопараметрическое семейство определенных на Λ функций:

$$S_\beta(z) = f + \beta\varphi^2 \qquad \text{(IV, 39)}$$

Поверхности уровня этих функций (в пространстве Z)

$$S_\beta(z) = c \qquad \text{(IV, 40)}$$

представляют собой параболоиды. При $c = \nu$, где

$$\nu = \min_{z \in \Lambda} S_\beta(z) \qquad \text{(IV, 41)}$$

формула (IV, 40) определяет параболоид, соприкасающийся с границей множества Λ (рис. 17). Пересечение соприкасающегося параболоида с осью Of определяется точкой $\nu\bar{e}$, причем

$$\nu = \min_{z \in \Lambda} S_\beta(z) \leqslant S_\beta[z(x^*)] = f(x^*) \qquad \text{(IV, 42)}$$

Если в выражении (IV, 42) $\nu < f(x^*)$, то можно улучшить нижнюю оценку ν для $f(x^*)$, увеличив β. Действительно, если $\beta' > \beta$, у параболоида

$$f + \beta'\varphi^2 = \nu \qquad \text{(IV, 43)}$$

(вершина в точке $\nu\bar{e}$) нет общих точек с множеством Λ, так как в противном случае [$\tilde{z} = (\tilde{f}, \tilde{\varphi}) \in \Lambda$ и удовлетворяет уравнению (IV, 43)] имело бы место равенство

$$S_\beta(\tilde{z}) = \tilde{f} + \beta\tilde{\varphi}^2 = \tilde{f} + \beta'\tilde{\varphi}^2 - (\beta' - \beta)\tilde{\varphi}^2 = \nu - (\beta' - \beta)\tilde{\varphi}^2 =$$
$$= \min_{z \in \Lambda} S_\beta(z) - (\beta' - \beta)\tilde{\varphi}^2 < \min_{z \in \Lambda} S_\beta(z)$$

что противоречит формуле (IV, 41). Тогда, принимая во внимание, что

$$S_{\beta'}(z) > S_\beta(z) \qquad \text{(IV, 44)}$$

получаем

$$\min_{z \in \Lambda} S_{\beta'}(z) = S_{\beta'}(z^*_{\beta'}) > \min_{z \in \Lambda} S_\beta(z) \qquad \text{(IV, 45)}$$

т. е.

$$\nu < \min_{z \in \Lambda} S_{\beta'}(z) = \nu' \leqslant f(x^*) \qquad \text{(IV, 46)}$$

Таким образом, двухуровневая вычислительная схема для метода штрафов предусматривает нахождение точной верхней грани $\sup\limits_{\beta' \in E^m} \nu'[S_{\beta'}(z)]$ на верхнем уровне и — минимизацию штрафной функции (IV, 39) на нижнем.

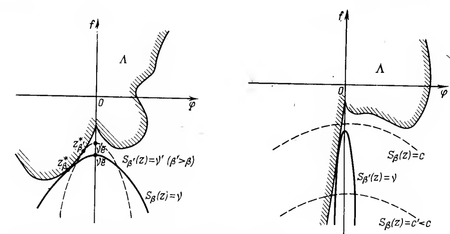


Рис. 17. Геометрическая интерпретация метода штрафов.

Рис. 18. Геометрическая интерпретация метода штрафов при f, неограниченной снизу.

В развернутом виде (в пространстве переменной x) функция (IV, 39) выглядит следующим образом:

$$P_\beta(x) = S_\beta[z(x)] = f(x) + \sum_{i=1}^{p} \alpha_i \varphi_i^2(x) + \sum_{j=1}^{q} \beta_i \psi_j^2(x) H[\psi_j(x)] \qquad \beta = (\alpha_i, \beta_j) \qquad \text{(IV, 47)}$$

Приведем алгоритм метода штрафов.

Алгоритм V.

1. Даны $x_0 \in E^n$, $\beta_0 \in E^m$, $\varepsilon > 0$. Положить $k = 0$.
2. Найти точку x_{k+1} минимума $P_{\beta_k}(x)$ функции (IV, 47).
3. Увеличить штрафные коэффициенты: $\beta_{k+1} = K\beta_k$ (обычно $K = 10$).
4. Если $|x_{k+1} - x_k| < \varepsilon$, то остановиться ($x_{k+1}$ — оптимальная точка), в противном случае положить $k = k + 1$ и перейти к шагу 2.

Условие остановки алгоритма (шаг 4) может быть дополнено или заменено условием (IV, 38).

Метод штрафов является более универсальным, чем метод множителей Лагранжа, применение которого возможно лишь при «стесняющем» условии — выпуклости множества Λ. В то же время применение метода штрафов имеет свои недостатки. Коэффициенты β, в общем случае, должны неограниченно возрастать, что приводит к «овражному» характеру функции $P_\beta(x)$ и, следовательно, существенно усложняет выполнение ее минимизации. Кроме того, если множество Λ не является ограниченным, точнее, если $f(x)$, $x \in \Omega$ не ограничена снизу, то выбор начальных величин β может оказаться проблематичным: при небольших β может случиться, что

$$\min_{x \in \Omega} P_\beta(x) = -\infty,$$

а при больших — сильно выраженная «овражность» $P_\beta(x)$ может оказаться непреодолимой при выполнении минимизации этой функции. Подобную ситуацию иллюстрирует график, приведенный на рис. 18. От этих недостатков свободен метод уровней, в котором не обязателен неограниченный рост коэффициентов в штрафных членах.

Следует отметить, что составная функция (IV, 36) метода уровней выглядит «сложнее»: если в функциях (IV, 7), (IV, 47), используемых соответственно в методах множителей Лагранжа и штрафных функций, критерий $f(x)$ содержится в «неискаженном» виде, то функция (IV, 36) метода уровней нелинейна относительно f.

Метод, использующий модифицированную функцию Лагранжа

Вернемся к задаче (IV, 1), (IV, 3). Обоснованное применение метода множителей Лагранжа возможно лишь при условии выпуклости множества Λ, выполнения которого, в общем случае, трудно ожидать и проверка которого при решении реальных задач практически неосуществима. Здесь будет рассмотрено обобщение метода множителей Лагранжа на случай невыпуклого множества Λ, представляющее собой своеобразный синтез этого метода с методом штрафов при конечной величине «штрафного» коэффициента [79]. В этом разделе по-прежнему будем считать множество Λ замкнутым, предполагая, что в точке x^* выполнено условие регулярности отображения $\varphi(x)$ [80, с. 74—75]. Более того, будем считать, что границу множества Λ в окрестности точки $z(x^*)$ можно аппроксимировать формой:

$$f - f(x^*) = a\varphi + \frac{1}{2}\varphi^{\mathrm{T}} A\varphi \qquad \text{(IV, 48)}$$

обладающей наименьшим собственным значением $\xi_{\min}$ матрицы A: $\xi_{\min} > -\infty$. Если $\xi_{\min} > 0$, то множество Λ локально выпукло в окрестности z^*, если же $\xi_{\min} \leqslant 0$, то существует конечное $\gamma > |\xi_{\min}|$, для которого множество Λ' в пространстве Z' является локально выпуклым в точке $z'^* = [f(x^*),\ 0_\Psi]$ при условии, что $z' = (f',\ \varphi)$, где

$$f' = f + \gamma\varphi^2 \qquad \text{(IV, 49)}$$

Таким образом, в случае невыпуклого (в точке z^*) множества Λ метод множителей Лагранжа может быть применен к функции (IV, 49), обладающей конечной величиной γ [79]:

$$f(x^*) = \Phi'_\gamma(x^*,\ \lambda^*) = \max_{\lambda} \min_{x \in \Omega} \Phi'_\gamma(x,\ \lambda), \qquad \text{(IV, 50)}$$

где

$$\Phi'_\gamma(x,\ \lambda) = f(x) - \lambda^{\mathrm{T}}\varphi(x) + \gamma\varphi^2(x) \qquad \text{(IV, 51)}$$

есть так называемая, модифицированная функция Лагранжа. В пространстве Z множество

$$\Phi'_\gamma(x,\ \lambda) = \min_{x \in \Omega} \Phi'_\gamma(x,\ \lambda) \qquad \text{(IV, 52)}$$

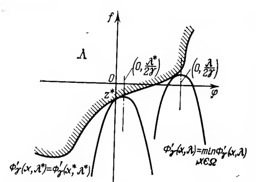


Рис. 19. Геометрическая интерпретация метода модифицированной функции Лагранжа.

представляет собой соприкасающийся с границей множества Λ параболоид, ось которого смещена относительно Of. На рис. 19 он показан справа. Координаты точки пересечения оси параболоида, показанной штрих-пунктирной прямой, с плоскостью $(0, \varphi)$ даны в скобках. В средней части рисунка изображен «оптимальный» параболоид, соприкасающийся с множеством Λ в точке z^*. Подчеркнем, что коэффициент γ в «штрафном» члене $\gamma\varphi^2$ выражения (IV, 49) — конечная величина, в то время как в методе штрафов, в общем случае, $\gamma \to +\infty$.

С учетом выражения (IV, 50) и выпуклости Λ' о характере двухуровневой вычислительной схемы в методе с модифицированной функцией Лагранжа может быть сказано то же самое, что и в методе множителей (см. Метод множителей Лагранжа). В частности, может быть использована рекуррентная формула (IV, 27) для вычисления последовательных значений λ. Запишем ее в следующем виде:

$$\lambda_{i+1} = \lambda_i - (UV^{-1}U^\tau)^{-1}\,\varphi\,(x_{i+1}) = \lambda_i -$$
$$- \left\{U\left[\nabla^2 f - \lambda^\tau\nabla^2\varphi + 2\gamma\,(\varphi^\tau\nabla^2\varphi + U^\tau U)\right]^{-1} U^\tau\right\}^{-1} \varphi\,(x_{i+1}) =$$
$$= \lambda_i - 2\gamma\left\{U\left[W + U^\tau U\right]^{-1} U^\tau\right\}^{-1} \varphi\,(x_{i+1}) \qquad \text{(IV, 53)}$$

где

$$W = \frac{1}{2\gamma}\left[\nabla^2 f - (\lambda - 2\gamma\varphi)^\tau\,\nabla^2\varphi\right] \qquad \text{(IV, 54)}$$

Для достаточно больших значений γ ($\gamma \gg 1$) членами второго порядка можно пренебречь. Находим

$$\lambda_{i+1} \approx \lambda_i - 2\gamma\left\{U\,(U^\tau U)^{-1}\,U^\tau\right\}^{-1} \varphi\,(x_{i+1}) \qquad \text{(IV, 55)}$$

Матрица в фигурных скобках выражения (IV, 55) представляет собой проекционный оператор [сравните с формулой (III, 38)], переводящий векторы из пространства E^p в подпространство, натянутое на векторы столбцов $\partial\varphi_j/\partial x$ матрицы U^τ. Так как, по предположению векторы $\partial\varphi_i/\partial x$, $i = 1, \ldots, p$ линейно-независимы, т. е. образуют базис в пространстве E^p, упомянутый проекционный оператор является тождественным преобразованием $[U\,(U^\tau U)^{-1}\,U^\tau = I_p]$ пространства E^p в себя и формула (IV, 55) принимает вид:

$$\lambda_{i+1} = \lambda_i - 2\gamma\varphi\,(x_{i+1}) \qquad \text{(IV, 56)}$$

Эта формула, представляющая собой «предельную» (для $\gamma \gg 1$) форму метода Ньютона для поиска максимума по λ функции

Φ'_γ $[x(\lambda), \lambda]$, находит применение в алгоритме минимизации с модифицированной функцией Лагранжа.

Отметим, что эту формулу можно получить с помощью простых, но менее строгих соображений. В точке минимума $x(\lambda_i)$ модифицированной функции Лагранжа $\Phi'_\gamma(x, \lambda_i)$ выполнено необходимое условие первого порядка

$$\frac{\partial \Phi'_\gamma [x(\lambda_i),\ \lambda_i]}{\partial x} = \frac{\partial f}{\partial x} - \left(\frac{\partial \varphi}{\partial x}\right)^{\text{т}} \lambda_i + 2\gamma \left(\frac{\partial \varphi}{\partial x}\right)^{\text{т}} \varphi \bigg|_{x=x(\lambda_i)} = 0 \qquad \text{(IV, 57)}$$

Этому условию можно придать вид необходимого условия Лагранжа (относительно λ) для исходной задачи с соответствующим образом подобранными $\bar{\lambda}$:

$$\frac{\partial \Phi[x(\lambda_i),\ \bar{\lambda}]}{\partial x} = \frac{\partial f}{\partial x} - \left(\frac{\partial \varphi}{\partial x}\right)^{\text{т}} \bar{\lambda} \bigg|_{x=x(\lambda_i)} = \qquad \text{(IV, 58)}$$

$$= \frac{\partial f}{\partial x} - \left(\frac{\partial \varphi}{\partial x}\right)^{\text{т}} (\lambda_i - 2\gamma\varphi) \bigg|_{x=x(\lambda_i)} = 0$$

$$\bar{\lambda} = \lambda_i - 2\gamma\varphi\,[x(\lambda_i)]$$

и принять эти значения множителей для последующей минимизации $\Phi'_\gamma(x, \lambda)$.

При больших значениях штрафного коэффициента $\gamma \gg 1$ в точке минимума $x_\gamma(\lambda)$ функции $\Phi'_\gamma(x, \lambda)$ в соответствии с (IV, 57) выполняется $\varphi(x_\gamma(\lambda)) \approx 0$ (исходя из предположения ограниченности первых производных функций f и φ_i). Таким образом, минимизация функции $\Phi'_\gamma(x, \lambda)$ при фиксированном значении λ с последующим его изменением по второй из формул (IV, 58) представляет собой итерационную процедуру решения системы приближенных уравнений необходимых условий оптимальности (уравнения Лагранжа)

$$\frac{\partial f}{\partial x} - \left(\frac{\partial \varphi}{\partial x}\right)^{\text{т}} \lambda = 0 \qquad \varphi(x) \approx 0$$

в виде решения системы $\text{F}[x(\lambda)] = 0$ (первое из необходимых условий), в котором $x(\lambda)$ определяется вторым из необходимых условий; изложенный подход полезно сравнить с аналогичной процедурой, описанной в работе [3, с. 92—93].

Сформулируем алгоритм минимизации для метода с модифицированной функцией Лагранжа.

Алгоритм VI.

1. Даны $x_0 \in E^n$, $\lambda_0 \in E^p_-$, $\gamma_0 > 0$, $\gamma_{\max} > \gamma_0$, $\varepsilon > 0$. Положить $k=0$.
2. Найти точку x_{k+1} минимума функции $\Phi'_{\gamma_k}(x, \lambda_k)$.
3. Вычислить «новые» значения λ_{k+1} множителей Лагранжа по формуле (IV, 56) или (IV, 27) и величины штрафных коэффициентов $\gamma_{k+1} = K\gamma_{k+1}$ $(K > 1)$, $\gamma_{k+1} \leqslant \gamma_{\max}$.
4. Если $\left|\partial\Phi'_{\gamma_{k+1}}(x_{k+1},\ \lambda_{k+1})/\partial x\right| \leqslant \varepsilon$, то остановиться ($x_{k+1}$ — оптимальная точка), в противном случае положить $k = k+1$ и перейти к шагу 2. Шаг 4 в этом алгоритме аналогичен шагу 4 алгоритма III. Подход, примененный в этом разделе, можно использовать в задачах многокритериальной оптимизации (приложение 1).

Совместное применение методов условной минимизации и квазиньютоновских методов безусловной минимизации

Рассмотрим теоретически некоторые особенности совместного применения методов уровней и штрафов и квадратичных методов безусловной минимизации, ориентируясь на эффективность алгоритма оптимизации в целом. Результаты такого рассмотрения могут оказаться полезными при анализе практического применения этих алгоритмов в каждом конкретном случае.

Предположим, что рассматриваемая задача содержит ограничения лишь в форме равенств. Обратимся к методу уровней. В этом случае минимизируемая функция имеет вид (IV, 36) с $\beta_j = 0$, $j = 1, \ldots, q$.

Вычислим матрицу вторых производных этой функции

$$\nabla_x^2 F_{\mu,\,\alpha}(x) = [f(x) - \mu]\,\nabla^2 f(x) + \nabla f(x)\,[\nabla f(x)]^{\mathrm{T}} +$$

$$+ \sum_{i=1}^{p} \alpha_i \left\{\varphi_i(x)\,\nabla^2\varphi_i(x) + \nabla\varphi_i(x)\,[\nabla\varphi_i(x)]^{\mathrm{T}}\right\}$$

Если x^* — решение исходной задачи, то при $\mu = \mu^*\ [f(x^*)]$, учитывая, что $\varphi_i(x^*) = 0$; $i = 1, \ldots, p$, находим

$$\nabla_x^2 F_{\mu^*,\,\alpha}(x^*) = \nabla f(x)\,[\nabla f(x)]^{\mathrm{T}} + \sum_{i=1}^{p} \alpha_i \nabla\varphi_i(x)\,[\nabla\varphi_i(x)]^{\mathrm{T}}\,\big|_{x=x^*}$$

Матрицы, входящие в правую часть этой формулы, имеют одинаковую форму представления: они получены в результате умножения вектора на свой транспонированный вектор. Ранг подобных матриц, очевидно, равен единице. Так как ранг суммы матриц не превосходит суммы рангов ее составляющих, то при $p < n - 1$ ранг матрицы $\nabla_x^2 F_{\mu^*,\,\alpha}(x^*)$ оказывается меньшим n, т. е. она является вырожденной. Если на нижнем уровне для минимизации функции $F_{\mu,\,\alpha}(x)$ применяется метод Ньютона [см. выражение (I, 43)], то в общем случае эффективность его для рассматриваемой ситуации значительно снижается [81, с. 79—86]: вместо квадратичной скорости сходимости можно гарантировать лишь линейную скорость, характерную для обычного градиентного метода. Следовательно в целом эффективность алгоритма метода уровней, используемого совместно с методом Ньютона для выполнения безусловной минимизации, должна снижаться по мере приближения значения параметра μ к μ^*. Отсюда следует также, что в общем случае метод уровней целесообразно применять лишь для локализации решения задачи на условный экстремум, в частности задавать начальные приближения для x и μ, достаточно близкие к x^*, μ^*, нецелесообразно. Последний из упомянутых моментов часто проявлялся при расчетах на ЭВМ с использованием на нижнем уровне других квадратичных методов безусловной минимизации.

Проанализируем теперь алгоритм метода штрафов. Характерной особенностью этого метода является неограниченный рост штрафных

коэффициентов α_i, $i = 1, \ldots, p$, в выражении для функции (IV, 47) (при ограничениях типа равенства $\beta_j = 0$, $j = 1, \ldots, q$). Обозначим эту функцию $P_\alpha(x)$. С практической точки зрения, чтобы получить достаточно точную оценку для x^*, приходится пользоваться довольно большими значениями α_i, $i = 1, \ldots, p$. При этом R — константа Липшица для матрицы вторых производных функции $P_\alpha(x)$ [при условии, что матрицы вторых производных функций $f(x)$ и $\varphi_i(\bar{x})$, $i = 1, \ldots, p$, удовлетворяют условию Липшица] увеличивается с ростом α_i, $i = 1, \ldots, p$. Оценки скорости сходимости метода Ньютона, используемого для минимизации функции $P_\alpha(x)$ на нижнем уровне, имеют вид [57, с. 71]:

$$|x_{k+1} - x^*| \leqslant \lambda_k |x_k - x^*|,$$

где

$$\lambda_k \leqslant \frac{R}{m} |x_k - x^*|$$

причем для $P_\alpha(x)$ предполагается выполненным условие сильной выпуклости

$$mz^{\text{т}}z \leqslant z^{\text{т}}\nabla_x^2 P_\alpha(x)\, z \leqslant Mz^{\text{т}}z \qquad M \geqslant m > 0$$

Окрестность x^*, в которой сходимость последовательности становится сверхлинейной, определяется условием

$$\frac{R}{m} |x_k - x^*| < 1$$

($\lambda_k < 1$), т. е. размер окрестности $r = |x_k - x^*|$ удовлетворяет неравенству $r < (m/R)$. Отсюда, чем больше значение константы Липшица R для $\nabla_x^2 P_\alpha(x)$, т. е. чем больше значения штрафных коэффициентов α_i, $i = 1, \ldots, p$, тем меньше размер окрестности, где скорость сходимости последовательности x_k к x^* становится сверхлинейной, т. е. где метод Ньютона обладает высокой эффективностью. Таким образом в целом с ростом α_i, $i = 1, \ldots, p$ эффективность метода штрафных функций при использовании метода Ньютона на нижнем уровне снижается.

Для преодоления указанных трудностей при решении задач на условный экстремум и уточнения положения экстремума целесообразно использовать комбинацию методов уровней и штрафов, первый — в начальной фазе оптимизации, второй — в заключительной. От этих недостатков свободен метод модифицированной функции Лагранжа, который находит в последнее время все более широкое применение.

Решение тестовых задач с ограничениями *

Результаты решения некоторых тестовых задач с ограничениями с помощью методов модифицированной функции Лагранжа (AL), уровней (MM) и штрафных функций (PEN) приведены в табл. 20, где Kl — число итераций на верхнем уровне, т. е. число изменений параметров составной функции; $\|\varphi_i\|$ и $\|\psi_j^+\|$ — нормы векторов ограничений типа равенства и неравенства в точке минимума x^*; λ_0

* Раздел написан совместно с Е. Ю. Койранской.

Таблица 20. **Результаты решения тестовых задач с ограничениями**

Номер тестовой задачи	Обозначение алгоритма минимизации: условной	Обозначение алгоритма минимизации: безусловной	K_f	K_p	K_l	f^*	$\lVert\varphi_i\rVert$	$\lVert\psi_j^+\rVert$	$\lambda_0\ (\mu_0)$	$\alpha_0-\alpha_{max}$	$\beta_0-\beta_{max}$
21 [164, с. 233]	AL	DFP	88	39	5	—44,4687	10^{-11}	10^{-10}	0	$1—10^5$	$10—10^5$
		SSVM	121	59	6	—44,4687	10^{-11}	10^{-11}	0	$1—10^6$	$10—10^6$
	MM	DFP	352	108	8	—44,4686	10^{-12}	0	—54,0	$1—10^6$	$1—10^6$
			290	100	7	—44,4686	10^{-12}	0	—54,0	$10—10^6$	$10—10^6$
			280	85	7	—44,4669	10^{-12}	0	—54,0	$10^2—10^6$	$10^2—10^6$
	PEN	DFP	243	97	7	—44,4687	10^{-7}	10^{-6}	—	$1—10^6$	$1—10^6$
			724	247	16	—44,4687	10^{-15}	10^{-15}	—	$1—10^{15}$	$1—10^{15}$
22 [164, с. 238]	AL	DFP	83	38	9	24,3062	—	10^{-5}	0	—	0,25
			100	40	6	24,3062	—	10^{-8}	0	—	0,25—4
			111	43	7	24,3062	—	10^{-8}	0	—	$10^{-1}—10^2$
			89	25	8	24,3062	—	10^{-8}	0	—	$10—10^2$
		SSVM	152	72	14	24,3062	—	10^{-5}	0	—	0,25
			133	61	6	24,3062	—	10^{-7}	0	—	0,25—4
	MM	DFP	184	72	7	24,3062	—	0	20,0	—	$10—10^3$
			129	57	11	24,3062	—	10^{-7}	20,0	—	10
			252	91	5	24,3063	—	0	0	—	$10—10^8$
			113	52	11	24,3062	—	10^{-7}	0	—	10

		SSVM	398	135	4	24,3077	—	0	0	—	$10—10^4$
	PEN	DFP SSVM	459 583	146 196	11 12	24,3062 24,3036	— —	10^{-6} 10^{-3}	— —	— —	$10—10^5$ $10—10^2$
23 [164, с. 239]	AL	DFP SSVM	170 280	64 118	8 8	130,490 130,490	— —	10^{-5} 10^{-5}	0 0	— —	0,25 0,25
	MM	DFP	1130 855	349 262	3 3	130,490 130,491	— —	0 0	0 100,0	— —	$10^2—10^8$ $10^2—10^8$
	PEN	DFP	1078	329	7	130,490	—	10^{-4}	—	—	$10—10^3$
24 * [98, с. 445]	AL	DFP SSVM	214 176	84 72	8 9	—47,761 —47,761	10^{-10} 10^{-9}	— —	0 0	1—64 1—64	— —
	PEN	DFP	614	226	11	—47,761	10^{-9}	—	—	$1—10^{10}$	—
25 [98, с. 465]	AL	DFP	377 889	119 288	10 11	32,3487 32,3487	— —	10^{-8} 10^{-10}	0 0	— —	10 $10—10^4$
	MM	DFP	2148	701	12	32,3487	—	10^{-10}	25,0	—	10^4

* При решении производилась замена переменных: $x_i = \bar{x}_i^2$

(или μ_0) — начальные значения множителей Лагранжа (параметра μ) в методе модифицированной функции Лагранжа (или методе уровней); α_0, β_0 — начальные значения весовых коэффициентов при штрафных членах. На нижнем уровне использовались методы безусловной минимизации DFP и SSVM (см. гл. III). В методах модифицированной функции Лагранжа и уровней вместе с изменением множителей λ (параметр μ) увеличивались значения весовых коэффициентов при штрафных членах в соответствующих составных функциях. Коэффициент менялся в пределах от 2 до 10. Значения весовых коэффициентов ограничивались сверху значениями α_{max}, β_{max}.

Из приведенных в табл. 20 результатов следует, что при прочих равных условиях, применение метода модифицированной функции Лагранжа оказывается предпочтительным (минимальное значение K_f). При заданной точности результата эффективность решения конкретной задачи существенным образом зависит от выбора начальных значений весовых коэффициентов α_0, β_0, коэффициента их изменения, начальных значений λ_0, μ_0.

Характеристика различных подходов к оптимизации химико-технологических систем *

В главе II было показано, что в зависимости от того, какие переменные в математической модели ХТС принимаются в качестве зависимых и независимых, возникают различные постановки задачи оптимизации ХТС. Для оценки той или иной постановки задачи удобно пользоваться следующими пятью характеристиками.

1. *Характер расчета критерия*; он зависит от того, необходима ли при заданных значениях независимых переменных итерационная процедура при вычислении значения критерия.

2. *Возможность использования алгоритмов условной или безусловной минимизации*. При прочих равных условиях алгоритмы безусловной минимизации существенно более просты. Поэтому преимущество имеют те подходы, которые обеспечат решение задачи с использованием только методов безусловной минимизации. В дальнейшем будем предполагать, что в качестве метода условной оптимизации используется метод штрафов.

3. *Число поисковых переменных*. Вообще говоря, с увеличением их числа эффективность квазиньютоновских методов падает.

4. *Простота учета ограничений на переменные*. Учет ограничений типа (I, 9) достаточно прост в рамках любых поисковых методов. В то же время учет ограничений типа (I, 10) может оказаться достаточно сложным, требующим трудоемкой процедуры.

5. *Возможность проведения параллельных вычислений*. Появление многопроцессорных ЭВМ, создание многомашинных комплексов в автоматизированных системах управления технологическими процессами требует разработки таких подходов, которые дали бы возможность эффективно использовать параллельно работающие процессоры или ЭВМ.

* Раздел написан совместно с Н. Н. Зиятдиновым.

Сравним рассматриваемые постановки задач оптимизации, исходя из перечисленных пяти характеристик.

Задача 1. Поиск в пространстве управлений [см. формулы (I, 64)—(I, 66)].

1. Для замкнутых схем расчет критерия достаточно трудоемок, поскольку требует решения системы нелинейных уравнений (I, 65), описывающей материальный и тепловой балансы ХТС.

2. Соотношения (I, 10) учитываются с помощью методов условной минимизации. Таким образом, в данном случае процедура оптимизации ХТС является трехуровневой (рис. 20): первый соответствует решению системы уравнений материального и теплового баланса схемы — системы (I, 65) — при фиксированных значениях поисковых переменных u, на втором переменные u изменяются в соответствии с каким-либо методом безусловной минимизации [возможно, учитывающим ограничения (I, 9)], третий соответствует изменению штрафного коэффициента в методе условной минимизации, который используется для выполнения ограничений (I, 10).

3. Число поисковых переменных равно $\bar{r}$ [см. соотношение (I, 52)].

4. В случае отсутствия ограничений (I, 10) для решения задачи 1 могут быть использованы методы безусловной минимизации, модифицированные для учета простейших ограничений (I, 9), и оптимизационная процедура будет состоять только из 1-го и 2-го уровней (рис. 20). В этом большое преимущество данного метода. Введение же ограничений (I, 10) существенно усложняет задачу, поскольку вследствие этого придется применять методы условной минимизации (появляется третий уровень в численной процедуре оптимизации).

5. Распараллеливание вычислений здесь возможно только на первом уровне. Система уравнений материального и теплового баланса схемы (I, 65) может решаться как последовательно, так и параллельно. В случае последовательного решения возможности параллельного счета очень ограничены. В случае параллельного решения системы уравнений (I, 65) можно организовать параллельный расчет блоков.

Задача 2. Поиск по некоторым управлениям при выполнении ограничений на выходные переменные при расчете схемы [см. соотношения (I, 71)—(I, 74)]. В качестве поисковых берут только часть управлений, а ограничения на выходные переменные ХТС автоматически удовлетворяются на каждом шаге оптимизационной процедуры.

1. Расчет критерия более трудоемок чем в задаче 1, поскольку требует решения системы уравнений материального и теплового балансов совместно с уравнениями ограничений на выходные переменные схемы — система (I, 72), (I, 73).

2. Поскольку соотношения (I, 10) автоматически выполняются на каждом шаге, для решения задачи 2 могут быть использованы методы безусловной минимизации, конечно, модифицированные для учета простейших ограничений (I, 9). В данном случае численная оптимизационная процедура будет двухуровневой (рис. 21) независимо от наличия или отсутствия ограничений (I, 10).

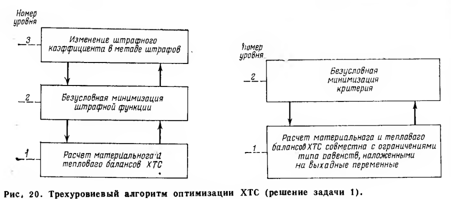


Рис. 20. Трехуровневый алгоритм оптимизации ХТС (решение задачи 1).

Рис. 21. Двухуровневый алгоритм оптимизации ХТС (решение задачи 2).

3. Число поисковых переменных несколько уменьшается по сравнению с предыдущим случаем. Оно равно $\bar{r} - \bar{s}$, где $\bar{s}$ определяется соотношением (I, 67).

4. Относительно распараллеливания вычислений здесь может быть сказано то же самое, что и в задаче 1.

Сравнивая подходы, использованные в задачах 1 и 2, можно отметить преимущество задачи 2, состоящее в том, что в ней используется только метод безусловной минимизации. Поэтому, если добавление системы (I, 73) к системе (I, 72) ненамного ухудшает скорость сходимости численной процедуры решения системы нелинейных уравнений на первом уровне (см. рис. 21), то этот подход может оказаться более эффективным. К недостаткам подхода, используемого в задаче 2, можно отнести следующее. Во время поиска переменные $\bar{u}$ могут принять такие значения, при которых система (I, 72), (I, 73) не будет иметь решения. Кроме того, при наличии ограничений на переменные и каждый раз когда одна из зависимых переменных будет выходить на ограничение, необходимо изменять совокупности зависимых $(\overline{u})$ и независимых $(\overline{\overline{u}})$ переменных.

Задача 3. Поиск в пространстве управлений и всех переменных состояния [см. формулы (I, 75)—(I, 76)].

1. Расчет критерия прост, поскольку как мы видели в этом случае фактически разрываются все потоки ХТС и на первом уровне приходится рассчитывать только левые части системы (I, 72), (I, 73).

2. Поскольку соотношения (I, 62), (I, 63) трактуются как ограничения типа равенств, в любом случае надо применять методы условной минимизации. Поэтому учет ограничений (I, 10) не меняет характера задачи, и сводится к добавлению штрафных членов в штрафной функции.

3. Число поисковых переменных $\bar{r} + m$ [см. соотношение (I, 52)] намного больше, чем в предыдущих случаях, поскольку как правило $m \gg \bar{r}$.

4. Все переменные x' являются независимыми, поэтому все левые части уравнений (I, 62) могут рассчитываться независимо для каж-

дого блока, что создает возможность параллельного счета на первом уровне.

Задача 4. Поиск в пространстве управлений и некоторых переменных состояния [см. формулы (I, 79)—(I, 81)].

1. Этот подход в основном применяют, чтобы добиться безытерационного расчета критерия. Выбор в качестве независимых всех входных переменных k-го блока эквивалентен разрыву всех его входных потоков. Поэтому в таком случае в качестве независимых переменных выбирают такие $x^{(k)}$, $(k = p_1, \ldots, p_q)$, при которых результирующая схема оказывается разомкнутой, а система уравнений (I, 80) допускает безытерационный расчет.

2. Учет ограничений (I, 81) проводится методами условной минимизации.

3. Условия распараллеливания вычислений будут целиком зависеть от того, какие из входных промежуточных переменных блоков схемы будут выбраны в качестве независимых переменных.

Сравнение постановок задач. Подход, использованный в задаче 3, в известном смысле противоположен подходам, использованным в задачах 1 и 2, основной недостаток которых обусловлен необходимостью проведения трудоемких процедур решения системы нелинейных уравнений на первом уровне и значительным усложнением процедуры решения задачи 1 при наличии ограничений (I, 10).

Основные недостатки использованного в задаче 3 подхода — это необходимость привлечения методов условной минимизации, существенное увеличение числа поисковых переменных и штрафных членов в модифицированном критерии. Поэтому в непосредственном виде этот подход в настоящее время практически не используется. Однако на его основе могут быть построены специальные методы оптимизации больших систем (см. гл. V).

Подход, использованный в задаче 4, является компромиссом между подходами, примененными в задачах 1 и 3, в нем сочетаются положительные стороны этих двух подходов, причем в некоторых случаях он может оказаться более эффективным. Действительно, пусть ограничения (I, 10) присутствуют и задача оптимизации ХТС сводится к задаче 1, для решения которой используется метод штрафов. Пусть выбор переменных $x^{(k)}$, $(k = p_1, \ldots, p_q)$ делает схему разомкнутой. Тогда, если суммарная размерность этих векторов мала по сравнению с $\bar{r}$, подход, использованный в задаче 4, может оказаться предпочтительным по сравнению с примененным в задаче 1, поскольку незначительно увеличивая размерность задачи, он делает расчет критерия безытерационным. Конечно, число штрафных членов в штрафной функции несколько увеличится (сравните D_4 и D_2). Однако мы исходим из предположения, что выполняется следующее свойство: если минимизируется штрафная функция, то добавление в нее небольшого числа новых штрафных членов, связанных с ограничениями типа (I, 56), ненамного ухудшает характеристики поиска. Хотя это правило нельзя доказать, более того, его можно и опровергнуть, построив специальные примеры, однако вычислительная

практика решения задач оптимизации ХТС показывает, что оно часто выполняется.

Оптимизация ХТС как многоуровневая процедура. Как мы видели, алгоритмы оптимизации ХТС, как правило, сводятся к многоуровневым процедурам (см. рис. 20, рис. 21). При построении эффективных алгоритмов многоуровневой оптимизации необходимо учитывать следующие правила.

Правило 1. Алгоритмы каждого уровня должны рассматриваться как взаимосвязанные части единого алгоритма оптимизации, которые должны быть тщательно согласованы один с другим.

Несколько усложнив работу алгоритмов одного уровня, можно существенно улучшить работу алгоритмов других уровней, что в целом приведет к более эффективной работе всей многоуровневой процедуры и, наоборот, улучшив работу алгоритмов одного уровня, можно усложнить работу алгоритмов других уровней, что приведет к менее эффективной работе всего алгоритма оптимизации.

Проиллюстрируем это положение двумя примерами.

За счет выбора поисковых переменных можно упростить задачу расчета схемы (первый уровень), сделав ее безытерационной (сведение задачи 1 к задаче 4), но существенно усложнить второй и третий уровень вследствие возрастания числа штрафных членов. И наоборот, можно усложнить алгоритм первого уровня, учтя в нем ограничения (I,10) (т. е. переходя к задаче 2), но существенно упростить вычисления на втором и третьем уровнях (третий уровень вообще может выпасть).

Другим примером может послужить выбор шага, т. е. величины коэффициента α_i в соотношении (I, 39) при линейном поиске в методе безусловной минимизации, т. е. на втором уровне (см. рис. 20). При применении методов безусловной оптимизации справедливо следующее: чем больше шаг вдоль направления, тем лучше. В том случае, когда первый уровень (расчет схемы) является безытерационным (задача 4), это справедливо и для многоуровневых процедур. В случае, когда первый уровень (расчет схемы) является итерационным (задача 1 для замкнутой схемы), это правило, вообще говоря, неверно. Действительно, при увеличении шага вдоль поискового направления действуют следующие противоположно направленные тенденции. С одной стороны увеличение шага вдоль направления дает хорошие результаты, поскольку уменьшается число итераций на втором уровне, но с другой стороны, увеличение шага ухудшает начальное приближение при решении системы (I, 65), что может привести к увеличению числа итераций на первом уровне. (При очень большом шаге квазиньютоновский метод на этом уровне вообще может перестать сходиться.) Должен существовать некоторый компромисс, при котором шаг вдоль направления будет наилучшим с точки зрения общего числа итераций на первом и втором уровнях.

Таким образом, критерием оценки эффективности процедуры оптимизации ХТС должно быть общее число k_s расчетов левых частей системы (I, 65) в случае задачи 1 или системы (I, 72), (I, 73) в случае задачи 2.

$$k_s = \sum_{i=1}^{k_f} l_i$$

где l_i — число расчетов левых частей этих систем при i-том вычислении критерия. Поскольку расчет левых частей системы (I, 72), (I, 73) эквивалентен однократному расчету соответствующей разомкнутой схемы, величина k_s равна общему числу однократных расчетов соответствующей разомкнутой схемы.

Интересно отметить, что в случае решения системы (I, 65) на первом уровне при слабой зависимости числа итераций от начального приближения может оказаться целесообразным вместо многократного решения системы (I, 65) на данном направлении p_i поиска безусловного минимума, осуществляемого на втором уровне, провести однократное решение системы $n + 1$-го уравнения:

$$\psi'(x, u_i + \alpha p_i) = 0 \qquad \frac{d\overline{F}}{dp_i} = 0$$

с $(n + 1)$ неизвестным — вектором x и скаляром α (здесь $\overline{F}$ — критерий оптимизации; p_i — i-тое направление поиска безусловной оптимизации; u_i — точка смены направления на $(i - 1)$-м направлении.

Перейдем теперь к формулированию 2-го правила. В многоуровневой процедуре j-му шагу k-го уровня соответствует некоторая итерационная процедура $(k - 1)$-го уровня. Так, j-му шагу алгоритма 2-го уровня (см. рис. 20) алгоритма безусловной минимизации соответствует итерационная процедура решения системы нелинейных уравнений (I, 65) расчета стационарного режима ХТС. Итерационному шагу 3-го уровня соответствует итерационная процедура безусловной минимизации 2-го уровня. Для простоты изложения итерационную процедуру $(k - 1)$-го уровня, соответствующую j-му шагу k-го уровня, будем называть j-той итерационной процедурой k-го уровня. Второе правило может быть записано следующим образом.

Правило 2. При построении начального приближения j-той итерационной процедуры k-го уровня необходимо максимально использовать информацию, полученную при проведении j-ой итерационной процедуры k-го уровня.

Рассмотрим алгоритм первого уровня (см. рис. 20). Ранее было показано, что при $i + 1$-м решении систем нелинейных уравнений стационарного режима ХТС квазиньютоновским методом может быть использована информация о решении и матрице Якоби, полученная при i-том решении системы нелинейных уравнений. При этом есть надежда, что такой прием окажется успешным, поскольку критерий как функция независимых переменных обычно является пологой функцией (кривизна ее мала), и шаг на втором уровне изменяет управления не на очень большую величину, т. е. выполняется условие (II, 196). При решении систем нелинейных уравнений 1-го уровня (см. рис. 20) методом Ньютона начальное приближение для $(i + 1)$-го решения систем нелинейных уравнений также может быть построено с использованием предыдущей информации (II, 200).

Рассмотрим теперь второй уровень алгоритма (см. рис. 20), соответствующий безусловной минимизации штрафной функции

$$F^{(k)} = f + \beta^{(k)}\varphi^2$$

при фиксированном значении штрафного коэффициента $\beta^{(k)}$. При этом для простоты мы предполагаем, что в задаче (IV, 1), (IV, 3), (IV, 4) имеется только одно ограничение типа равенства. Значение коэффициента $\beta^{(k)}$ назначается на третьем уровне. Обозначим через H^{*j} предельное значение матрицы H_i, полученной при минимизации $F^{(j)}$ каким-либо квазиньютоновским методом минимизации. Поскольку H^{*j} будет близко к обратному гессиану функции $F^{(j)}$, естественно при минимизации функции $F^{(j+1)}$ для $\beta^{(j+1)} = k\beta^{(j)}$, ($k > 1$) взять в качестве H_0 матрицу H^{*j}. Аналогичным образом можно поступить и при использовании метода уровней и модифицированного метода множителей Лагранжа. Однако, при значительном увеличении штрафного коэффициента здесь может быть допущена большая погрешность. Поэтому рассмотрим другой прием. Пусть будет аппроксимироваться сам гессиан, и при $\beta^{(j)}$ проведена оптимизация функции $F^{(j)}$, в результате которой получено значение предельной матрицы B^{*j}. Выполняется соотношение

$$\nabla^2 F^{(j)} = \nabla^2 f + \beta^{(j)} \nabla^2 \varphi^2 \qquad \text{(IV, 59)}$$

Надо оценить гессиан $\nabla^2 F^{(j+1)}$ в той же точке

$$\nabla^2 F^{(j+1)} = \nabla^2 f + \beta^{(j+1)} \nabla^2 \varphi^2$$

Подставляя в это соотношение $\nabla^2 f$ из соотношения (IV, 59), а также заменяя $\nabla^2 F^{(j)}$ близкой к нему матрицей B^{*j}, получим

$$\nabla^2 F^{(j+1)} = B^{*j} + (\beta^{(j+1)} - \beta^{(j)}) \nabla^2 \varphi^2$$

Воспользовавшись аппроксимацией [11, с. 133] для $\nabla^2 \varphi^2$, можем найти $\nabla^2 F^{(j+1)}$ и взять ее в качестве начального приближения для матрицы B при минимизации функции $F^{(j+1)}$. Рассмотрим теперь выбор шага, т. е. коэффициента α_i в уравнении (I, 39) при линейном поиске в методе безусловной минимизации, для случая, соответствующего задаче 4. Поскольку при этом первый уровень безытерационный, на величину α не накладывается ограничений со стороны первого уровня и он должен выбираться из условия, чтобы число вычислений целевой функции было минимально. На каждом направлении происходит линейный поиск точки, удовлетворяющей условиям (I, 47). Он начинается с начального шага α_i^0. Выбор этой величины существенно влияет на число шагов на данном направлении, а, следовательно, и на общее число шагов. Естественно предположить, что на двух соседних направлениях i-том и $(i + 1)$-м расстояния $\| x_{i+1} - x_i \|$ и $\| x_i - x_{i+1} \|$ будут близки, поэтому начальный шаг на $(i + 1)$-м направлении разумно выбирать из условия $\alpha_{i+1}^0 = k\alpha_i^*$, где множитель k близок к единице и выбирается на опыте по данным вычислительного эксперимента.

Остановимся теперь на задаче определения частных производных критерия при применении первого подхода; в качестве примера возьмем схему, приведенную на рис. 5. Для простоты будем считать, что в каждом блоке имеется только одна управляющая переменная. Тогда при применении разностного метода вычисления производных потребуется $N + 1$ итерационных процедур расчета замкнутой

схемы. Опишем более экономный способ вычисления производных. Пусть итерируемыми переменными при расчете схемы (см. рис. 20, уровень 1) будут входные переменные блока *2*. Обозначим их через x. Тогда расчет сводится к решению системы (II, 6), где компоненты вектора F являются выходными переменными блока *1* разомкнутой схемы (см. рис. 6). Критерий Φ можно считать сложной функцией переменных

$$\Phi = \Phi(x, u)$$

где x — неявная функция переменных u, определяемая соотношением (II, 6). Продифференцируем Φ по переменным u

$$\frac{\partial \Phi}{\partial u} = \frac{\partial \Phi}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial \Phi}{\partial u} \qquad \text{(IV, 60)}$$

$$\left(I_n - \frac{\partial \Phi}{\partial u} \right) \frac{\partial x}{\partial u} = - \frac{\partial F}{\partial u} \qquad \text{(IV, 61)}$$

Решив систему линейных уравнений (IV, 61), найдем производные $\partial x/\partial u$. Подставив их в выражение (IV, 60), найдем искомые производные. Для их определения необходимо знать производные $\partial \Phi/\partial x$, $\partial \Phi/\partial u$, $\partial F/\partial x$, $\partial F/\partial u$. С точки зрения схемной интерпретации эти величины ($\partial \Phi/\partial u$, $\partial \Phi/\partial x$) представляют собой производные критерия по управлениям и входным переменным разомкнутой схемы (см. рис. 6) соответственно, а $\partial F/\partial x$, $\partial F/\partial u$ — производные выходных переменных блока *1* по входным и управляющим переменным разомкнутой схемы (см. рис. 6).

Итак, для определения производных критерия оптимизации замкнутой схемы необходимо рассчитать частные производные ряда величин разомкнутой схемы. Определение этих величин не требует проведения итерационных процедур. В этом состоит основное преимущество даиного подхода. Кроме того, при вычислении производиых в разомкиутой схеме можно воспользоваться «зоиами влияиия» [3, с. 136], что может также существеино сократить число вычислений. Правда, использоваиие этого подхода требует решеиия системы лииейиых уравиений. Покажем, что используя информацию, полученную на первом уровне (см. рис. 20), можно еще более повысить эффективиость этого метода. Будем исходить из предположеиия, что для решения системы (II, 6) на первом уровие (см. рис. 20) используется квазиньютоновский метод QNM. Обозначим через H^* предельное зиачеиие матрицы H_i [см. соотиошение (II, 101)]. Матрица H_i аппроксимирует обратиую матрицу Якоби системы (II, 6), в пределе можио ожидать, что матрица H_i стремится к обратиой матрице Якоби этой системы, т. е. что будет выполняться равеиство

$$H^* = \left(I_n - \frac{\partial F}{\partial x} \right)^{-1}$$

Отсюда, воспользовавшись выражением (IV, 61), получим

$$\frac{\partial x}{\partial u} = -H^* \frac{\partial F}{\partial u}$$

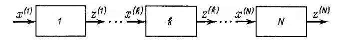


Рис. 22. Схема последовательности блоков.

Вычисление производных с помощью этой формулы не требует во-первых, решения системы линейных уравнений, а во-вторых, вычисления производных $\partial F/\partial \overline{\overline{x}}$, что существенно сократит объем вычислений. Таким образом, обеспечение хорошей сходимости матрицы H_i к матрице Якоби системы (II, 6) дает экономный способ определения искомых производных. Более того, подставив матрицу $\partial x/\partial u$ в формулу (II, 200), можно получить начальное приближение для решения системы (II, 6) при $u = u_{i+1}$.

Выбор поисковых переменных. Как мы видели, каждая из рассмотренных задач имеет свои отрицательные и положительные стороны. Естественно поставить вопрос о том, в каких случаях целесообразно применять ту или иную из них. Такие рекомендации можно дать только для отдельных случаев:

	Номер задачи, которую целесообразно использовать
Схема разомкнута, отсутствуют ограничения (I,10)	1
Схема разомкнута, имеются ограничения (I,10), производные критерия вычисляются аналитически *	1 или 2
Схема замкнута, большинство блоков охвачено обратными связями	4

* Как будет показано далее при разностном вычислении производных более эффективным может оказаться четвертый подход

Поскольку выбор той или иной задачи связан с выбором поисковых (независимых) переменных, поставленный вопрос фактически сводится к выбору наилучшей совокупности поисковых переменных. Проблема близка к проблеме выбора множества разрываемых потоков в задаче расчета стационарного режима схемы [3, с. 33]: и в том и в другом случае речь идет о выборе наилучшей совокупности итерируемых переменных. Задача выбора оптимальной совокупности поисковых переменных при оптимизации ХТС чрезвычайно сложна. Сложность ее связана с трудностью формулирования критерия, который позволял бы оценивать тот или иной набор поисковых переменных без решения задачи оптимизации ХТС. Наилучшим критерием, конечно, является время, затраченное на решение задачи, но его можно определить только, решив задачу оптимизации. Поэтому цель состоит в том, чтобы косвенным путем оценить это время. Строго поставить и сформулировать эту задачу трудно. Придется пользоваться некоторыми эвристическими правилами, экспериментальными фактами, которые помогут сформулировать эту задачу хотя бы и не очень строго.

В виде примера рассмотрим задачу оптимизации последовательности блоков (рис. 22). В данном случае $m_k = n_k = n$, $(k = \overline{1, N})$, $s_1 = n$, $s_k = 0$, $(k = \overline{2, N})$, $g_k = 0$, $(k = \overline{1, N-1})$, $g_N = n$, $r_k = r$, $(k = \overline{1, N})$. Соотношения связи (I, 2) имеют простой вид:

$$x_i^{(k)} - y_i^{(k-1)} = 0 \qquad i = \overline{1, n} \qquad k = \overline{2, N} \qquad \text{(IV, 62)}$$

В случае сведения задачи оптимизации к задаче 1, когда поисковыми являются управления всех блоков схемы, на втором уровне (рис. 20) будет решаться следующая задача

$$\min_{u^{(k)}} \left[F + \beta \sum_{j=1}^{\bar{g}_N} \left(z_j^{(N)} - a_j^{(N)} \right)^2 \right] \qquad k = \overline{1, N} \qquad \beta \gg 1 \qquad (IV, 63)$$

где $z_j^{(N)}$ определяются последовательным расчетом всех блоков при заданных значениях $u^{(k)}$, $(k = \overline{1, N})$.

Рассмотрим теперь постановку, приводящую к задаче 4. В качестве поисковых переменных наряду с управлениями $u^{(k)}$, $(k = \overline{1, N})$ выберем переменные

$$x_i^{(k)} \qquad (i = \overline{1, n};\ k = p_1, \ldots, p_q) \qquad (IV, 64)$$

Тогда на втором уровне будет решаться следующая задача:

$$\min_{u^{(k)},\ x^{(s)}} \left(F + \beta_1 \sum_{j=1}^{\bar{g}_N} \left(z_j^{(N)} - a_j \right)^2 + \beta_2 \sum_{i=1}^{n} \sum_{k=p_1}^{p_q} \left(x_i^{(k)} - z_i^{(k-1)} \right) \right)$$

$$s = p_1, \ldots, p_q \qquad k = \overline{1, N} \qquad (IV, 65)$$

Число поисковых переменных в данном случае будет равно $m = r_1 + \ldots + r_N + qn$.

Будем считать, что частные производные вычисляются с помощью разностей. Поставим задачу определения «наилучшей» совокупности переменных (IV, 64). Обозначим через $\tau_k = \tau$, $(k = \overline{1, N})$, T_1, T_2 соответственно, время расчета k-го блока, схемы и градиента критерия. Ясно, что $T_1 = \tau N$. Сделаем следующие допущения.

1. Число поисковых направлений при решении задачи оптимизации пропорционально размерности задачи, т. е. равно γm, где γ — некоторый коэффициент пропорциональности.

2. Число δ вычислений критерия на каждом поисковом направлении постоянно. Вычислительный опыт показывает, что величина δ действительно является достаточно стабильной. Так, для алгоритма, используемого в нашей практике, она лежит в пределах от 4 до 6.

3. Выполняется свойство штрафных функций (см. с. 129). Используя первые два допущения, получим грубую оценку времени T, затрачиваемого на решение задачи оптимизации:

$$T = \delta \gamma m T_1 + \gamma m T_2 = \gamma m (\delta T_1 + T_2)$$

Величину T будем использовать в качестве критерия для выбора переменных (II, 64). Для вычисления частных производных используем «зоны влияния» [3, с. 136], применение которых позволяет существенно сократить число вычислений. Напомним, что k-тый блок последовательности (см. рис. 22) имеет «зону влияния», состоящую из блоков k, $k + 1$, ..., N. Совокупность блоков с номерами от p_k до $p_{k+1} - 1$ будем называть k-тым участком.

Рассмотрим частный случай, когда последовательность блоков (см. рис. 22) разбивается на участки с равным числом блоков, т. е. $p_{k+1} - p_k = N/(q + 1)$. Новыми выходными переменными схемы будут $z^{(i)}$, $(i = p_1 - 1, \ldots, p_q - 1)$. Время вычисления частных производных критерия по управлениям и входным переменным x^{p_k} $(k + 1)$-го участка $(k = \overline{1, q})$ равно

$$\frac{\tau n N}{q + 1} + \frac{1}{2} \frac{rn}{q + 1} \left(\frac{N}{q + 1} + 1 \right)$$

Отсюда общее время T равно [101]

$$T = \tau \gamma (rN + nq) \left[\delta N + nN + 0{,}5 rN \left(\frac{N}{q + 1} + 1 \right) \right]$$

Итак, задача выбора оптимальной совокупности дополнительных переменных (IV, 64) свелась к минимизации функции T по целочисленной переменной q. Из этой

формулы ясен эффект введения переменных (IV, 64). Действительно, если $Nr \gg nq$, то с увеличением q член, стоящий в первых скобках, увеличивается незначительно, а третье слагаемое во вторых скобках уменьшается примерно пропорционально q. Следовательно, при достаточно больших N увеличение q может привести как к уменьшению величины, стоящей в этой скобке, так и общей величины T.

Использование этого подхода к задаче оптимизации последовательности реакторов идеального смешения [11, с. 50] показало его эффективность. На этом примере ясно видна также польза введения дополнительных поисковых переменных для распараллеливания вычислений в случае использования многопроцессорных ЭВМ или многомашинных комплексов. Благодаря последовательной структуре схемы (см. рис. 22) здесь может эффективно использоваться только одна ЭВМ. Введение же дополнительных поисковых переменных позволяет параллельно обрабатывать отдельные участки на нескольких ЭВМ.

Оптимизация процесса Вильямса—Отто

В главе II было дано описание химико-технологического процесса Вильямса—Отто и приведена его математическая модель. Задача оптимизации заключается в том, чтобы выбрать такой режим проведения процесса, при котором годовой доход, выраженный в процентах к сумме вложений в производство, оказывается максимальным. Рассмотрим составляющие дохода:

Объем сбыта (за 1 ч)	$0{,}3F_P + 0{,}0068F_D$
Стоимость сырья (за 1 ч)	$0{,}02F_A + 0{,}03F_B$
Затраты на обработку отходов (за 1 ч)	$0{,}01F_G$
Водо-, паро-, энергоснабжение (за 1 год) . . .	$2{,}22F_R$
Издержки на сбыт, администрацию и научные исследования, % объема сбыта	12,4
Затраты на амортизацию и ремонт, % суммы капитальных вложений/год	10
Капитальные затраты	$600V\rho$

Продолжительность работы производства в течение года составляет в среднем 8400 ч. Критерий оптимизации (максимизации) принимает вид:

$$f(x) = 100\,[8400\,(0{,}3F_P + 0{,}0068F_D - 0{,}02F_A - 0{,}03F_B - 0{,}01F_G) - 0{,}124\cdot 8400\,(0{,}3F_P + 0{,}0068F_D) - 60V\rho - 2{,}22F_R]/600V\rho \qquad (IV, 66)$$

В дополнение к обычно используемым варьируемым переменным F_A, F_B, T и коэффициенту рецикла α введем V — объем реактора. Оптимальный режим должен удовлетворять заданным ограничениям на температуру T внутри реактора и производительность процесса F_P по целевому продукту P:

$$580 \leqslant T \leqslant 680 \qquad (IV, 67)$$

$$0 \leqslant F_P \leqslant 4763 \qquad (IV, 68)$$

Выражения для F_P, F_G и F_D, необходимые для расчета $f(x)$ по формуле (IV, 66), имеют вид:

$$F_P = F_{RP} - 0{,}1F_{RE} \qquad (IV, 69)$$

$$F_G = 1{,}5k_3F_{RC}F_{RP}V\rho/F_R^2 \qquad (IV, 70)$$

$$F_D = (1-\alpha)\,(F_R - F_G - F_P) \qquad (IV, 71)$$

Наряду с приведенной выше постановкой задачи оптимизации, где ограничения на производительность по продукту P имеют вид неравенств (IV, 68), рассматривался другой вариант этой задачи, в котором фиксирована производительность по P:

$$F_P = c = \text{const} \qquad \text{(IV, 72)}$$

Значение постоянной c принималось равным 4763 (для сравнения задач в первой и второй постановках по трудоемкости их решения), а в некоторых случаях равным 4300. При проведении решения ограничение (IV, 72) учитывалось путем введения соответствующего штрафного члена в функцию условной минимизации. Учет уравнений модели выполнялся двумя способами. В первом из них переменные F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE} включались в список варьируемых переменных задачи оптимизации (в дополнение к F_A, F_B, T, V, α). В таком случае уравнения модели (II, 124)—(II, 126) при постановке задачи оптимизации интерпретировались как ограничения типа равенства (на эти переменные) и учитывались механизмом условной оптимизации [см. задачу 4, (I, 79)—(I, 81)]. При втором способе совокупность уравнений модели (II, 124)—(II, 126) трактовалась как система нелинейных уравнений относительно F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE}, решение которой выполнялось для каждого фиксированного набора значений варьируемых переменных F_A, F_B, T, V, α, вычисленных в процессе оптимизации [см. задачу 1, выражения (I, 64)—(I, 66)]. Решение системы уравнений модели выполнялось как различными методами [QNM и Broyden, см. формулы (II, 121), (II, 49)], так и простой итерацией. Использовались модификации QNM1 и Broyden 1 с уменьшением квадрата нормы левых частей системы на каждом из направлений движения. В качестве начального приближения к решению F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE} системы для текущего набора значений варьируемых переменных F_A, F_B, T, V, α принимались значения F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE}, найденные для предыдущих значений F_A, F_B, T, V, α. Кроме того, в модификациях QNM1H, Broyden 1H в качестве начальной аппроксимации H_0 обратного якобиана при данных значениях F_A, F_B, T, V, α использовалась матрица H, определенная в результате решения системы уравнений модели для предыдущих значений F_A, F_B, T, V, α. В качестве алгоритма оптимизации применялся метод модифицированной функции Лагранжа в сочетании с методом DFP на нижнем уровне и разностным способом расчета производных. В табл. 21 приведены результаты решения задачи оптимизации с условиями (IV, 68), а в табл. 22 — для задачи с фиксированной производительностью (IV, 72).

Каждый результат представлен тремя числами: K_f, K_p, K_t (K_t — число обращений к расчету «разомкнутой» схемы, включая обращения для расчета производных с помощью разностей).

В качестве начальных приближений для переменных F_A, F_B, T, V, α и F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE} использовались значения [39], соответствующие $V = 60$, приведенные в табл. 3 (см. начальные точки 1, 2, 3). Решение задачи удалось получить не для всех вариантов, так

Таблица 21. **Результаты оптимизации процесса Вильямса—Отто (производительиость по продукту P ие задаиа)**

Номер начальной точки	Результаты оптимизации					
	уравнення моделн трактуются как ограничения типа равенств	уравнення модели трактуются как снстемы нелинейных уравнений				
		простая итерация	QNM1	Broyden1	QNM1H	Broyden 1H
1	19891 5800 77892	114 27 23375	324 36 5249	—	220 24 1551	—
2	3116 992 13037	138 32 27522	137 22 2160	132 19 1990	180 22 4331	—
3	2009 602 8030	217 45 28642	109 20 1750	132 27 2390	92 19 1357	88 17 723

Таблица 22. **Результаты оптимизации процесса Вильямса—Отто при заданной производительиости по продукту P ($F_P = \text{const}$)**

Номер начальной точки	Результаты оптимизации									
	уравиеиия модели трактуются как ограничения типа равенств		уравнения модели трактуются как снстемы нелинейных уравнеиий							
			простая итерация	QNM1		Broyden 1		QNM1H	Broyden 1H	
	$F_P=4763$	$F_P=4363$	$F_P=4763$	$F_P=4763$	$F_P=4300$	$F_P=4763$	$F_P=4300$	$F_P=4763$	$F_P=4763$	$F_P=4300$
1	3736 1233 16063	5064 1588 20945	172 26 25533	135 37 2691	155 43 3067	158 41 16224	156 47 3087	136 33 1294	146 34 1294	144 41 1507
2	3610 1168 15291	2090 686 8931	141 18 18806	122 35 2440	143 36 2711	143 36 10726	128 36 2549	130 35 1243	143 39 1438	— — —
3	2215 717 9386	2513 800 10514	103 23 16131	86 18 1511	89 20 1659	76 18 1409	89 20 1597	89 21 736	89 19 883	— — —

как в некоторых случаях метод модифицированной функции Лагранжа, будучи методом внешней точки, приводил к появлению (в процессе решения) таких значений F_A, F_B, T, V, α при которых методы QNM или (и) Broyden решения системы (II, 124)—(II, 126) не сходились, либо давали отрицательные значения величин F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE}. Появление таких ситуаций частично удалось преодолеть введением соответствующих замен переменных, например, $F'_{RA} = \overline{F}^2_{RA}$ для предотвращения отрицательных значений F_{RA} при

решении системы уравнений, или переход от α к $\bar{\alpha}$ с помощью функции arctg, обеспечивающий выполнение неравенств $0 < \alpha < 1$ при проведении оптимизации.

В оставшихся нерешенными вариантах неудача может быть объяснена либо отсутствием решения системы (II, 124)—(II, 126) в какой-либо промежуточной точке F_A, F_B, T, V, α оптимизации, либо достаточной «удаленностью» промежуточной точки оптимизации от предыдущей (несмотря на предусмотренное в алгоритме оптимизации ограничение на величину $\bar{\rho}$ ($\leqslant \bar{\rho}_{max}$) начального шага по направлению поиска), в результате чего начальное приближение по F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE}, являющееся решением системы для предшествующей точки в пространстве варьируемых переменных, оказывается неудовлетворительным для данной (текущей) точки F_A, F_B, T, V, α и алгоритмы QNM и Broyden не «срабатывают».

Анализ результатов показывает, что наибольшей эффективностью (наименьшее K_t — число расчетов «разомкнутой» схемы) обладает «традиционный» подход — решение системы уравнений (II, 124)—(II, 126) относительно переменных F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE} для каждой точки F_A, F_B, T, V, α, определяемой в процессе оптимизации, при условии, что решение системы выполняется достаточно сильными методами (QNM, Broyden). При этом использование модификаций алгоритма: QNM1H, Broyden 1H дает наилучшие результаты. Для вариантов с начальными точками 2 и 3 (см. табл. 22) и при использовании метода QNM1H имелось несколько результатов, для которых $\bar{\rho}_{max} = 10^{-5} - 10^{-1}$; при изменении $\bar{\rho}_{max}$ в этих границах с ростом $\bar{\rho}_{max}$ значение K_t сначала уменьшалось, а затем увеличивалось; в таблице приведен «оптимальный» по $\bar{\rho}_{max}$ ($\bar{\rho}_{max} = 0{,}001$) результат.

Применение простой итерации для решения уравнений модели (II, 124)—(II, 126) оказывается нецелесообразным и дает наихудшие результаты. Если уравнения модели (II, 124)—(II, 126) учесть в алгоритме оптимизации в качестве ограничений на переменные F_{RA}, F_{RB}, F_{RC}, F_{RP}, F_{RE}, то по числу вычислений этот случай будет промежуточным. Сопоставление результатов, приведенных в первых столбцах табл. 21, 22, показывает, что добавление еще одного ограничения ($F_P = \text{const}$) к уже имеющимся не приводит к существенному увеличению значений K_f, K_p, K_t при таком способе решения задачи.

В заключение отметим, что при традиционном подходе к решению задачи оптимизации среднее число обращений к расчету разомкнутой схемы, приходящееся на одно вычисление критерия оптимизации (включая разностные оценки производных), колеблется в пределах: для метода простой итерации — от 70 до 100, а для квазиньютоновских методов — от 4 до 9.

Решение задачи оптимизации процесса разделения изомеров октана

Описание схемы процесса разделения изомеров октана (см. рис. 1), его математическая модель и постановка задачи оптимизации были приведены в главе I. В схеме процесса присутствует циркулирующий

Таблица 23. **Результаты оптимизации * для различных Δx_i**

$\frac{\Delta x_i}{x_i}$	K_f	K_p	K_l	$f*$	$\|\varphi_i\|$	$\|\psi_j^+\|$	λ_0	$\alpha_0-\alpha_{max}$	$\beta_0-\beta_{max}$
10^{-3}	1087	186	10	25,6	10^{-5}	0	0	$1-10^5$	$1-10^5$
10^{-4}	842	189	6	25,5	10^{-4}	0	0	$1-10^5$	$1-10^5$
10^{-5}	671	153	7	25,6	10^{-6}	10^{-6}	0	$1-10^5$	$1-10^5$
10^{-6}	1395	343	6	25,6	10^{-5}	10^{-5}	0	$1-10^5$	$1-10^5$
10^{-7}	1262	308	6	25,6	10^{-6}	0	0	$1-10^5$	$1-10^5$
10^{-8}	1230	262	6	25,6	10^{-6}	10^{-6}	0	$1-10^5$	$1-10^5$
10^{-9}	1583	409	8	591,3	10^{-2}	0	0	$1-10^5$	$1-10^5$
10^{-10}	916	199	6	25,6	10^{-6}	0	0	$1-10^5$	$1-10^5$

* Коэффициент при y_G в выражении (I,29) был принят равным ≈16,967.

поток I_R, возвращаемый на вход схемы с выхода блока изомеризации. Рецикл можно учесть двумя способами: на уровне расчета схемы при итерациях по x_1 [см. задачу 1, выражения (I, 64)—(I, 66)] и при оптимизации, рассматривая его как ограничение типа равенства на разрываемую переменную x_1 [см. задачу 4, выражения (I, 79)—(I, 81)]. При решении был применен второй способ. Оптимизация проводилась с применением методов последовательной безусловной минимизации: метода модифицированной функции Лагранжа (AL) и штрафных функций (PEN), на нижнем уровне которых использовались квазиньютоновские алгоритмы DFP, SSVM. Расчет производных выполнялся разностным способом [см. выражение (I, 49)]. В процессе оптимизации для удержания значений варьируемых переменных x_i (напомним, что x_i — коэффициенты разделения газовых потоков) между нулем и единицей применялись замены переменных с использованием функции arctg. Функции, участвующие в постановке задачи оптимизации, наиболее чувствительны (в окрестности x^*) к изменению x_1, x_3, x_7. В связи с этим для повышения стабильности получаемых результатов применялось преобразование сжатия по осям x_2, x_4, x_5, x_6, что можно сравнить с процедурой [11, с. 82—83]. В табл. 23 приведены результаты решения рассматриваемой задачи

Таблица 24. **Результаты оптимизации процесса получения изомеров октана** $\left(\frac{\Delta x_i}{x_i}=10^{-5}\right)$

Обозначение метода минимизации: условной	Обозначение метода минимизации: безусловной	K_f	K_p	K_l	$f*$	$\|\varphi_i\|$	$\|\psi_j^+\|$	λ_0	$\alpha_0-\alpha_{max}$	$\beta_0-\beta_{max}$
AL	DFP	671	153	7	25,6	10^{-6}	10^{-6}	0	$1-10^5$	$1-10^5$
AL	SSVM	1285	337	8	593,9	10^{-4}	0	0	$1-10^5$	$1-10^5$
PEN	BFGS	2721	559	5	591,7	10^{-3}	0	0	$1-10^5$	$1-10^5$
PEN	DFP	767	180	8	25,6	10^{-6}	10^{-5}	—	$1-10^9$	$1-10^9$

методом модифицированной функции Лагранжа с использованием DFP, относящиеся к выбору приращений Δx_i варьируемых переменных для расчета градиента разностным способом, а в табл. 24 — результаты оптимизации с применением других методов решения. При этом в случае использования метода множителей Лагранжа при изменении λ весовые коэффициенты в штрафных членах увеличивались вдвое, а при методе штрафных функций — на порядок. Найденное оптимальное значение критерия составляет $f^* = 25{,}6$ % при $x^* = (1834{,}22;\ 0;\ 0{,}63533;\ 1;\ 0;\ 1;\ 0{,}751795)$.

Расчет оптимальных стационарных режимов контактных узлов сернокислотного производства

Одной из основных химических реакций в процессе получения серной кислоты является каталитическое окисление диоксида серы в триоксид

$$SO_2 + \frac{1}{2} O_2 \rightleftarrows SO_3 \qquad (IV, 73)$$

В промышленных условиях эта реакция осуществляется в контактных аппаратах, представляющих собой многослойный каталитический реактор с встроенными между слоями и выносными теплообменниками, предназначенными для отвода реакционного тепла. Основное применение в сернокислотной промышленности получили схемы контактных узлов, работающих по методу одинарного (одностадийного) контактирования (рис. 23) и по методу двойного контактирования и двойной абсорбции (рис. 24). Последний метод предполагает организацию двухстадийного контактирования. На рис. 24 представлена схема (3 + 1), первая стадия которой включает первые три слоя катализатора, а вторая — последний слой в реакторе. Каждая из стадий контактирования завершается абсорбцией SO_3. Разделение процесса окисления на две стадии с последующей абсорбцией SO_3 способствует увеличению скорости реакции (IV,73) на заключительной (второй) стадии вследствие значительного снижения эффекта торможения реакции продуктом SO_3, что позволяет достичь более высокой степени превращения SO_2 в SO_3 по сравнению с получаемой при одностадийных схемах контактирования.

На рис. 23 приведена схема весьма распространенного на химических заводах пятислойного контактного аппарата с поддувом свежего газа после первого слоя катализатора. Поток сернистого газа G, поступающего на окисление, последовательно нагревается во внешнем теплообменнике и в теплообменнике после второго

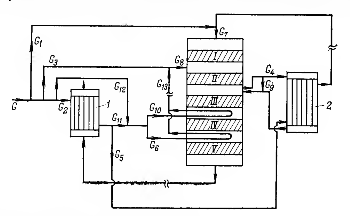


Рис. 23. Схема контактного узла окисления SO_2:
1, 2 — внешний и выносной теплообменники; *I—V* — слои катализатора.

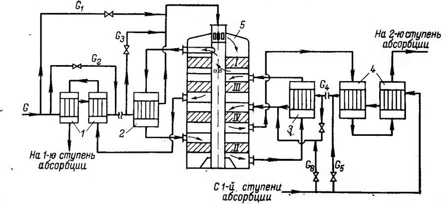


Рис. 24. Схема контактного узла окисления SO_2 по методу двойного контактирования и двойной абсорбции:

1—4 — теплообменники; *5* — реактор; *I—IV* — слои катализатора.

слоя катализатора до температуры, обеспечивающей устойчивое протекание реакции (IV,73), а затем поступает в первый слой катализатора, где (IV,73) проводится адиабатически (без отвода тепла) до достижения температуры, близкой к равновесной, определяющей прекращение реакции окисления. Снижение температуры окислившегося в первом слое газа, необходимое для дальнейшего проведения окисления SO_2, достигается добавкой (поддувом) относительно холодного диоксида серы (поток G_8) в смеситель, установленный в аппарате. Охлажденный газ поступает во второй слой катализатора, где протекает последующее окисление диоксида серы, сопровождаемое выделением тепла и повышением температуры до величины, близкой к равновесной. Теплообменник, куда поступает газ, после второго слоя катализатора, снижает температуру реакционной смеси, которая затем подается в третий слой, и т. д. Промежуточные теплообменники выполняют двойную функцию: охлаждают окислившийся в соответствующем слое катализатора газ и нагревают газ, подаваемый в первый слой катализатора (теплообменник после второго слоя) и на поддув перед вторым слоем (теплообменники после третьего и четвертого слоев). С целью управления температурным режимом контактного узла газовая нагрузка G аппарата перераспределяется между теплообменниками (потоки G_5, G_8, G_{10}), добавляется поток G_1 холодного газа, подаваемый на вход первого слоя катализатора, а часть G_{12} холодного газа байпасируется мимо внешнего теплообменника.

Таким образом, контактный узел сернокислотного производства представляет собой сложную химико-технологическую систему, характеризующуюся наличием взаимовлияющих параметров и обратных тепловых потоков.

Математическая модель контактного узла [82, 83] описывает элементы схемы и связи между ними.

Слой катализатора. В i-том ($i = 1, \ldots, 5$) слое катализатора процесс окисления диоксида серы в триоксид, протекающий по реакции (IV, 73), описывается уравнением [84]

$$\frac{dz^{(i)}}{d\tau} = \frac{k^{(i)}}{a} \cdot \frac{b - \frac{az^{(i)}}{2}}{1 - \frac{az^{(i)}}{2}} \cdot \frac{1 - z^{(i)}}{1 - \frac{1}{4} z^{(i)}} \times$$

$$\times \left[1 - \frac{(z^{(i)})^2}{(K_p^{(i)})^2 (1 - z^{(i)})^2} \cdot \frac{1 - \frac{az^{(i)}}{2}}{b - \frac{az^{(i)}}{2}} \right] \qquad 0 \leqslant \tau \leqslant \bar{\tau}^{(i)} \qquad \text{(IV, 74)}$$

совместно с уравнением

$$t^{(i)}(\tau) = t^{(i)}(0) + \lambda\,[z^{(i)}(\tau) - z^{(i)}(0)] \qquad \text{(IV, 75)}$$

где

$$k^{(i)} = k_0^{(i)} \exp\left[-\frac{E^{(i)}}{R}\left(\frac{1}{T^{(i)}} - \frac{1}{T^*}\right)\right] \qquad \text{(IV, 76)}$$

Здесь $z^{(i)}$ — степень контактирования, т. е. доля окислившегося диоксида серы от общего его содержания в исходном газе; τ — текущее время контакта; $k^{(i)}$ — константа скорости реакции (IV, 73); $k_0^{(i)}$ — константа, характеризующая катализатор; $E^{(i)}$ — энергия активации; R — газовая постоянная; $T^{(i)}$ — некоторая характерная для данного катализатора температура; $t^{(i)}$ — текущее значение температуры; $t^{(i)}$ (0) — температура газового потока на входе в слой катализатора; λ — коэффициент адиабатического разогрева; $z^{(i)}$ (0) — начальная степень контактирования (на входе в слой); a — концентрация сернистого ангидрида в исходном газе (в долях единицы); b — концентрация кислорода в исходном газе (в долях единицы); $K_p^{(i)}$ — константа равновесия реакции (IV, 73).

Равновесное значение $z_p^{(i)}$ степени контактирования определяется соотношением

$$z_p^{(i)} = \frac{K_p^{(i)}}{K_p^{(i)} + \sqrt{\left(1 - \frac{1}{2}\,az_p^{(i)}\right)\Big/\left(b - \frac{1}{2}\,az_p^{(i)}\right)}} \qquad \text{(IV, 77)}$$

Если степень контактирования, определяемая выражением (II, 74), достигает своего равновесного значения с заданной точностью при некотором $\tau = \tilde{\tau} < \bar{\tau}^{(i)}$:

$$z^{(i)}(\tilde{\tau}) \approx z_p^{(i)}$$

то процесс интегрирования уравнений (IV, 74), (IV, 75) обрывается, поскольку при этом считается, что дальнейшего изменения (в пределах заданной точности) степени контактирования $z^{(i)}$ и температуры $t^{(i)}$ в слое катализатора не происходит

$$z^{(i)}(\tau) = z^{(i)}(\tilde{\tau}) = z_p^{(i)} \qquad t^{(i)}(\tau) = t^{(i)}(\tilde{\tau}) = t_p^{(i)} \qquad \tilde{\tau} \leqslant \tau \leqslant \bar{\tau}^{(i)} \qquad \text{(IV, 78)}$$

Теплообменник. Математическое описание теплообменников содержит уравнения материального и теплового балансов. Уравнения теплового баланса теплообменника могут быть записаны в виде

$$Q = G_{\text{о}}c_{\text{о}}\,(t_{\text{о. н}} - t_{\text{о. к}})\,\eta \qquad Q = G_{\text{в}}c_{\text{в}}\,(t_{\text{в. к}} - t_{\text{в. н}})\,\eta \qquad Q = kF\,\Delta t_{\text{ср. л}}\varphi \qquad \text{(IV, 79)}$$

где Q — количество тепла, отдаваемое «горячим» газом; G — поток теплоносителя (индексы «о», «в» относятся соответственно к величинам, характеризующим потоки отдающие и воспринимающие тепло); c — теплоемкость газа; t — температура газового потока (индексы «н», «к» относятся соответственно к начальному и конечному ее значениям); η (≈ 1) — коэффициент, учитывающий тепловые потери в окружающую среду; k — коэффициент теплопередачи; F — по-

верхность теплообмена; $\Delta t_{\text{ср. л}}$ — среднелогарифмическая разность температур; φ — коэффициент, учитывающий отклонение движения теплоносителей от идеального противотока. Из уравнений (IV, 79) следует:

$$t_{\text{в. к}} = t_{\text{в. н}} + (t_{\text{о. н}} - t_{\text{в. н}})\, N\, \frac{\exp[2M(1-N)] - 1}{\exp[2M(1-N)] - N} \qquad \text{(IV, 80)}$$

$$t_{\text{о. к}} = t_{\text{о. н}} - (t_{\text{о. н}} - t_{\text{в. н}})\, \frac{\exp[2M(1-N)] - 1}{\exp[2M(1-N)] - N} \qquad \text{(IV, 81)}$$

где

$$N = \frac{G_{\text{о}} c_{\text{о}}}{G_{\text{в}} c_{\text{в}}} \qquad M = \frac{kF\varphi}{2G_{\text{о}} c_{\text{о}}} \qquad \text{(IV, 82)}$$

При известных режимных ($t_{\text{о. н}}$, $t_{\text{в. н}}$, $G_{\text{о}}$, $G_{\text{в}}$) и конструктивных (число труб, диаметр, длина) параметрах теплообменника формулы (IV, 80), (IV, 81) позволяют вычислить температуры охлаждаемого и нагреваемого газовых потоков на выходе из теплообменника.

Соотношения связи между слоями катализатора и теплообменниками определяются из уравнений материального (сохранения общего количества вещества) и теплового балансов в соответствии со структурой схемы. Так например, согласно схеме контактного аппарата (см. рис. 23) величина G_7 газового потока на входе в первый слой катализатора определяется соотношением

$$G_7 = G_5 + G_1 \qquad \text{(IV, 83)}$$

а уравнение теплового баланса

$$G_7 t^{(1)}(0) = G_5 t_{\text{вк}}^{(2)} + G_1 t_{\text{хол}} \qquad \text{(IV, 84)}$$

позволяет определить температуру газа на входе в первый слой катализатора

$$t^{(1)}(0) = \frac{G_5 t_{\text{вк}}^{(2)} + G_1 t_{\text{хол}}}{G_7} \qquad \text{(IV, 85(}$$

и т. д.

Здесь и далее приняты следующие обозначения: $t^{(i)}(0)$, $t^{(i)}(\bar{\tau}^{(i)})$, $z^{(i)}(0)$, $z^{(i)}(\bar{\tau}^{(i)})$ — начальные и конечные значения температуры и степени контактирования для i-го ($i = 1, 2, \ldots, 5$) слоя катализатора; $t_{\text{о. н}}^{(i)}$, $t_{\text{о. к}}^{(i)}$, $t_{\text{в. н}}^{(i)}$, $t_{\text{в. к}}^{(i)}$ — начальные (индекс «н») и конечные (индекс — «к») температуры отдающего (индекс «о») и воспринимающего (индекс «в») тепло потоков в теплообменнике после i-го (i=2, 3, 4) слоя (принятые обозначения при $i = 1$ относятся к температурам газовых потоков для внешнего теплообменника); $t_{\text{хол}}$ — температура газового потока G на входе в контактный аппарат.

При окислении диоксида серы в триоксид объем газа уменьшается, т. е. реакция (IV, 73) идет с уменьшением объема. Поэтому количество газовой смеси, поступающей в теплообменники из слоев катализатора, определяется по формуле

$$G_{\text{о}}^{(i)} = G\left[1 - \frac{1}{2} a z^{(i)}(\bar{\tau}^{(i)})\right] \qquad i = 2,\ 3,\ 4, \qquad G_{\text{о}}^{(1)} = G\left[1 - \frac{1}{2} a z^{(5)}(\bar{\tau}^{(5)})\right] \qquad \text{(IV, 86)}$$

где $G_o^{(i)}$ — поток газа, отдающего тепло в i-том теплообменнике. Количества газа в потоках, воспринимающих тепло, и их начальные температуры также определяются из уравнений материального и теплового балансов. Так например, равные температуры на входе в теплообменники после 3-го и 4-го слоев $t_{в.\,н}^{(3)}$, $t_{в.\,н}^{(4)}$ определяются из следующего соотношения теплового баланса:

$$(G_6 + G_{10})\, t_{в.\,н}^{(3)} = G_{11} t_{в.\,к}^{(1)} + G_{12} t_{хол} \qquad \text{(IV, 87)}$$

Математическая модель контактного аппарата, содержащая математическое описание слоев катализатора (IV, 74)—(IV, 78), теплообменников (IV, 80)—(IV, 82) и связей между отдельными элементами контактного узла соотношения, аналогичные выражениям (IV, 83), (IV, 84), (IV, 87) позволяет рассчитывать стационарные режимы аппарата при различных значениях его конструктивных и технологических параметров.

Формулировка задачи оптимизации. В качестве критерия оптимизации стационарного режима контактного узла принята степень контактирования во всем аппарате $z^{(5)}\,(\bar{\tau}^{(5)})$. Определение оптимального режима, обеспечивающего максимально возможное значение $z^{(5)}\,(\bar{\tau}^{(5)})$, ведет к более полному использованию сырья и снижению выбросов непрореагировавшего сернистого ангидрида в окружающую среду. В качестве варьируемых переменных $\bar{\alpha}_k$ ($k = 1, \ldots, 6$) приняты расходы газа в узловых точках разделения потоков, выраженные в долях газовой нагрузки аппарата:

$$\bar{\alpha}_k = G_k/G \qquad k = 1, \ldots, 6 \qquad \text{(IV, 88)}$$

Таким образом, задача оптимизации стационарных режимов контактных узлов сернокислотного производства заключается в определении значений $\bar{\alpha}_k$, позволяющих реализовать такие значения температур $t^{(i)}\,(0)$ газа на входе в слои катализатора, которые обеспечили бы достижение максимальной степени контактирования во всем аппарате при существующих значениях газовой нагрузки G, концентрации a сернистого газа в исходной газовой смеси, состоянии $k_0^{(i)}$, катализатора и конструктивных параметрах Z слоев катализатора и теплообменников. Следовательно, максимизируемая функция имеет вид:

$$z^{(5)}\,(\bar{\tau}^{(5)}) = f\left(\bar{\alpha}_k,\ G,\ k_0^{(i)},\ Z\right) \qquad \text{(IV, 89)}$$

На варьируемые переменные $\bar{\alpha}_k$ наложены ограничения

$$(\bar{\alpha}_k)_{\min} \leqslant \bar{\alpha}_k \leqslant (\bar{\alpha}_k)_{\max} \qquad k = 1, \ldots, 6 \qquad \text{(IV, 90)}$$

и дополнительные ограничения (IV, 91), имеющие следующий смысл:

Ограничение	Описываемое условие	
$\bar{\alpha}_5 \leqslant \bar{\alpha}_2$	... $G_{11} \geqslant 0$	(IV, 91)
$\bar{\alpha}_1 + \bar{\alpha}_2 + \bar{\alpha}_3 \leqslant 1$	... $G_{12} \geqslant 0$	
$\bar{\alpha}_1 + \bar{\alpha}_3 + \bar{\alpha}_5 + \bar{\alpha}_6 \leqslant 1$	... $G_{10} \geqslant 0$	

Ранее было отмечено, что контактные узлы сернокислотного производства (см. рис. 23, 24) содержат обратные связи по теплу между реакционной смесью и исходным газом, т. е. представляют собой замкнутые химико-технологические системы. Как показано в работах [85, 86], наличие в схемах контактных узлов обратных тепловых потоков может привести к появлению неустойчивых режимов при определенных значениях параметров. При этом условия баланса по веществу и теплу в разрывах обратных потоков, выполнения которых обычно достигают при проведении итерационного расчета схемы относительно переменных в «разрывах», целесообразно перенести на уровень оптимизации, рассматривая их как ограничения типа равенства и считая переменные в разрывах дополнительными варьируемыми переменными [см. задачу 4, выражения (I, 79)—(I, 81)]. Это позволяет в каждой точке расширенного пространства варьируемых переменных, полученной в процессе оптимизации, выполнять расчет лишь разомкнутой схемы, и, таким образом, избежать при выполнении вычислений появления нежелательных «нулевых» режимов и неоднократной проверки условий неустойчивости. Эти условия достаточно проверить лишь в конечной (оптимальной) точке. Таким образом, прием вынесения ограничений в критерий оптимизации (составную функцию), позволяет перейти к эквивалентной задаче оптимизации для разомкнутой схемы в расширенном пространстве варьируемых переменных.

Эквивалентная задача (впрочем, как и исходная) представляет собой задачу на условный экстремум, для решения которой использовалась условная оптимизация: метод уровней и метод модифицированной функции Лагранжа. Для выполнения безусловной минимизации составной функции (нижний уровень оптимизации) применялись методы квазиньютоновского типа — DFP, BFGS, SSVM [см. (III, 81), (III, 84)]. Расчет производных минимизируемой функции выполнялся как «аналитически» — с привлечением сопряженного процесса [3, с. 142], так и методом конечных разностей, что позволило провести сравнение результатов оптимизации по эффективности и точности решения *.

В принятой модели контактного узла каждый поток характеризуется тремя величинами, вычисленными в данном сечении потока: $\{\alpha, t, z\}$, где α — величина (объемный расход) потока, выраженная в долях газовой нагрузки на аппарат; t — температура; z — степень превращения SO_2 в SO_3. Слой катализатора обладает одним выходным потоком, параметры z и t, которого изменяются в соответствии с уравнениями (IV, 74), (IV, 75), и одним выходным, при этом относительная величина потока считается неизменной в процессе интегрирования. Смесители газовых потоков обладают соответственно двумя входными и одним выходным делителем газовых потоков — одним входным и двумя выходными потоками. Для каждого из указанных трех элементов схемы зависимость параметров выходного потока от

* Разработка версии программы для ЭВМ, включающей сопряженный процесс, выполнена Коноплевой А. Ю.

входных значений (уравнения модели) может быть задана в векторном виде:

$$y = f(x, u) \qquad (IV, 92)$$

где y — выходные значения параметров $\{\alpha, t, z\}$; x — входные переменные; u — варьируемые переменные (они появляются лишь в уравнениях делителей и имеют смысл коэффициентов деления потоков). Теплообменник обладает двумя входными потоками (отдающий и воспринимающий тепло) и, соответственно, двумя выходными. Согласно уравнениям (IV, 80)—(IV, 82) связь между его входными (x) и выходными (y) параметрами записывается в виде неявной функции:

$$\varphi(x, y) = 0 \qquad (IV, 93)$$

так как при расчете коэффициента теплопередачи используются значения теплофизических свойств газов, вычисленных при некоторых «средних» значениях температуры.

Принятый здесь подход к расчету процесса контактирования в последовательных слоях катализатора заключается в расчете процесса, протекающего в одном «эквивалентном» слое. Последний представляется в виде совокупности подслоев, в которой объединены данные слои контактной массы со своими характеристиками. При расчете «эквивалентного» слоя изменение параметра α в каждом подслое и в промежуточных потоках между его подслоями не учитывается, так как эффект изменения объема реакционной смеси представлен кинетическим уравнением (IV, 74), а влияние промежуточных теплообменников проявляется в скачкообразном изменении температуры на границах раздела подслоев. Уравнения «сопряженных» блоков для элементов (IV, 92) имеют вид:

$$\mu_j = \sum_i \frac{\partial f_i}{\partial x_j} \lambda_i \qquad (IV, 94)$$

где λ, μ — сопряженные переменные, представляющие соответственно, вход и выход «сопряженного» элемента. Для каждого слоя частные производные его выходных величин (T, z) по входным параметрам определяются, как результат интегрирования соответствующей системы уравнений в вариациях, которое выполняется одновременно с интегрированием основного уравнения (IV, 74) вместе с уравнением (IV, 75). Аналогичный вид имеют уравнения блоков, «сопряженных» с системой уравнений (IV, 93); при этом частные производные правых частей блоков (IV, 94) вычисляются дифференцированием неявной функции (IV, 93).

Расчет сопряженного процесса, содержащего блоки (IV, 94), выполняется в «обратном» порядке по отношению к порядку расчета элементов основного процесса. Необходимые «начальные» значения для сопряженных переменных λ (на входе «сопряженного» процесса и в «разрывах», соответствующих «разрывам» основного процесса) определяются частными производными минимизируемой (составной) функции по соответствующим переменным основного процесса.

Некоторые из параметров $\{\alpha, t, z\}$ определенных потоков схемы зависят лишь от значений параметров входного потока контактного узла, которые предполагаются фиксированными. Это обстоятельство явно учитывалось при вычислении производных $\partial f_i/\partial x_j$ для уравнения (IV, 94); некоторые из них полагались тождественно равными нулю.

Заметим, что присутствие дополнительных технологических ограничений (например, на выходную температуру первого слоя), представляющих собой ограничения на промежуточные переменные системы, приводит к появлению соответствующего «штрафного» члена в составной функции. Расчет производных критерия, содержащего явную зависимость от промежуточных переменных системы, связан в этом случае с несколько видоизмененной формой системы (IV, 94), описывающей сопряженный блок: в ее правую часть вводится свободный член — производная критерия по соответствующей промежуточной переменной.

Наконец, следует отметить, что в реальных условиях устойчивость режима контактного узла достигается путем определенного увеличения объема контактной массы в слоях контактного аппарата. Тем самым создается, так называемый, запас катализатора, причем условия неустойчивости процесса, о которых говорилось ранее, оказываются заведомо невыполненными [86]. Устойчивый ход процесса окисления зависит, главным образом, от величины запаса катализатора в первых двух слоях контактного аппарата и связан с поддержанием в этих слоях определенного режима контактирования, при котором значения степеней превращения диоксида серы в триоксид на выходе каждого из них оказываются практически равновесными. В этом случае условие устойчивости процесса может быть учтено в алгоритме оптимизации как ограничение на величину (коэффициент) запаса катализатора в данном слое, значения которой не могут быть менее заданной. Введем время контакта в слое — отношение объема контактной массы к объемной скорости газового потока через слой. Тогда коэффициент запаса слоя можно представить в виде

$$m = \bar{\tau}/\tau \qquad \text{(IV, 95)}$$

где $\bar{\tau}$ — полное время контакта в слое; τ — «момент» времени контакта, при котором величина степени превращения диоксида серы достигает (с точностью ε) своего равновесного значения, т. е.

$$\omega(x_0, \tau) = z_p - z(x_0, \tau) = \varepsilon \qquad \text{(IV, 96)}$$

где z_0 — определяется выражением (IV, 77), а $x_0 = \{\alpha, t, z\}$ — параметры газового потока на входе в слой. Принимая, что коэффициент запаса есть выходная переменная слоя, приведем выражения для его производных по входным параметрам данного слоя, необходимые по уравнению (IV, 94) для формирования сопряженного процесса.

В соответствии с соотношением (IV, 77) величина z_p является функцией константы равновесия K_p (при фиксированных a, b), известным образом зависящей лишь от температуры t в данном сечении слоя. В свою очередь t зависит от $z(x_0, \tau)$ и x_0 [см. выражения (IV, 75)]. Таким образом, $z_p = z_p(x_0, \tau)$. Соотношение (IV, 96,

Таблица 25. **Оптимизация контактного узла (см. рис. 23) при $G = 65\,000$ м³, $a_{SO_2} = 7,5$ %**

Подход к условной оптимизации	DFP	SSVM	BFGS
Метод модифицированной функции Лагранжа	97,91 % 192; 58,3	97,87 % 200; 73; 4	97,83 % 217; 57; 3
Метод «уровней»	97,90 % 208; 58; 2	97,90 % 199; 61; 2	97,84 % 198; 55; 2

определяет τ как неявную функцию от x_0, т. е. $\tau = \tau(x_0)$. Выполняя дифференцирование по x_0 тождества $\omega[x_0, \tau(x_0)] \equiv \varepsilon$, нетрудно выразить производные $\partial\tau/\partial x_0$ через $\partial\omega/\partial x_0$ и $\partial\omega/\partial\tau$, которые могут быть легко вычислены с помощью соотношения (IV, 77) и интегрирования системы уравнений в вариациях для уравнения (IV, 74). Искомые производные коэффициента запаса m по входным переменным слоя x_0 имеют вид

$$\frac{\partial m}{\partial x_0} = -\frac{\bar{\tau}}{\tau^2}\frac{\partial\tau}{\partial x_0} = -\frac{m}{\tau}\frac{\partial\tau}{\partial x_0}$$

В табл. 25 приведены результаты расчетов оптимального режима контактного узла одинарного контактирования (см. рис. 23) с применением сопряженного процесса. Каждый результат представлен величиной $z^{(5)}(\bar{\tau}^{(5)})$, выраженной в %, и тройкой чисел K_f, K_p, K_l. Подобные расчеты выполнялись ранее [11, с. 180] с применением разностной оценки производных (там же приведены все необходимые параметры контактного узла). Сравнение этих двух подходов показывает, что в данном случае использование сопряженного процесса для расчета градиента минимизируемой функции сокращает затраты времени ЭВМ при поиске оптимального режима в два раза. При этом точность определения оптимального режима в обоих случаях примерно одинакова.

Минимизация функций при наличии линейных ограничений

При решении задач оптимизации химико-технологических процессов очень часто ограничения на управляющие переменные являются линейными. Часто они имеют характер простых ограничений на максимальные и минимальные значения соответствующих управляющих переменных (I,9). В схемах, как правило, имеются делители потоков, на коэффициенты деления которых налагаются линейные ограничения вида (I,7). Особенно много таких ограничений будет в задачах синтеза при применении метода структурных параметров (см. гл. VI). Конечно, для решения задачи оптимизации с линейными ограничениями, можно использовать общие методы, разработанные для случая произвольных ограничений. Однако этот случай можно рассматривать отдельно по двум причинам. Первая из них состоит в том, что в задачах, где имеются только линейные ограничения, удается построить более эффективные алгоритмы, используя линейный характер ограничений. Вторая причина состоит в следующем. Математические модели отдельных аппаратов часто могут работать только в некоторой допустимой области. Скажем, если во время оптимизационной процедуры концентраций какой-либо компоненты на входе реактора примет

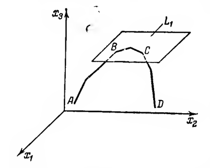


Рис. 25. Пример поисковой траектории.

отрицательное значение вследствие несовершенства применяемого метода оптимизации, то математическая модель, реактора, не рассчитанная на такой физически нереальный случай, может перестать работать. Причем часто допустимая область задается линейными ограничениями. В то же время при использовании общих методов во время поиска ограничения могут нарушаться, что может привести к прекращению поиска. Увеличивая штрафные коэффициенты, можно добиться уменьшения нарушения ограничений, но это приводит к ухудшению характеристик сходимости. В то же время при использовании методов, рассчитанных на учет линейных ограничений, удается достаточно точно удовлетворять их.

Рассмотрим методы, в которых аппроксимируется обратный гессиан. Итак, рассмотрим задачу:

$$\min_x f(x) \qquad \text{(IV, 97)}$$

$$n_i^{\mathrm{T}}x - b_i = 0 \qquad i = \overline{1,\ m} \qquad \left(n_i^{\mathrm{T}}n_i = 1\right) \qquad \text{(IV, 98)}$$

$$n_i^{\mathrm{T}}x - b_i \leqslant 0 \qquad i = m+1,\ \ldots,\ p \qquad \text{(IV, 99)}$$

В процессе поиска некоторые из ограничений — неравенства (IV,99) — могут стать строгими равенствами. Эти ограничения будем называть активными. Обозначим через L_q, $(q \geqslant m)$ линейное многообразие, образованное пересечением q гиперплоскостей, из которых m являются гиперплоскостями, описанными уравнениями (IV, 98), а $(q - m)$ — гиперплоскостями, соответствующими $(q - m)$ активным ограничениям (IV, 99). Совокупность m ограничений (IV, 98) и всех активных ограничений в данной точке списка будем называть б а з и с о м.

Во время поиска изображающая точка может двигаться в поисковом пространстве внутри некоторого многообразия L_q $(q \geqslant m)$, может входить на новые гиперплоскости, соответствующие активным ограничениям (увеличение базиса), или наоборот, сходить с некоторой гиперплоскости (уменьшение базиса). В качестве примера на рис. 25 изображена поисковая траектория в трехмерном пространстве при наличии одного ограничения типа неравенства (IV, 99), причем допустимая область находится ниже плоскости L_1. Поиск начинается в точке A. Участок траектории AB лежит в полном пространстве (базис пустой). В точке B поисковая траектория пересекает плоскость L_1 и в дальнейшем лежит в этой плоскости. Здесь уже имеется одно активное ограничение (базис состоит из уравнения плоскости L_1). В точке C поисковая траектория сходит с плоскости L_1 и далее участок траектории CD уже лежит в полном пространстве. Отсюда ясно, что любой алгоритм, осуществляемый в соответствии с формулами (I,39), (I,41), должен содержать следующие основные алгоритмы.

1. Алгоритм движения при неизменном базисе. Этот алгоритм в свою очередь должен состоять из трех частей. Первая часть — алгоритм определения матрицы в H_i в выражении (I,41), при неизменном базисе; вторая часть — определение шага вдоль поискового направления и третья часть — критерий схода с активного ограничения.

2. Алгоритм изменения матрицы H_i при увеличении числа активных ограничений.

3. Алгоритм изменения матрицы H_i при уменьшении числа активных ограничений.

Вначале выведем некоторые вспомогательные соотношения для случая, когда имеются ограничения только типа равенств

$$n_i^{\mathrm{T}}x - b_i = 0 \qquad i = \overline{1,\ q},\ q \geqslant m \qquad \text{(IV, 100)}$$

Пусть $N_q = \{n_1, \ldots, n_q\}$ $(n \times q)$-матрица, столбцы которой образованы векторами n_i, $(i = \overline{1, q})$. Будем считать, что ее ранг равен q. Алгоритм движения надо строить таким образом, чтобы поисковые направления p_i лежали внутри линейного многообразия L_q, образованного пересечением гиперплоскостей (IV, 100). Другими словами, должны выполняться соотношения

$$p_k^{\mathrm{T}} n_i = 0 \qquad k = 0,\ 1,\ \ldots \qquad i = \overline{1,\ q} \qquad \text{(IV, 101)}$$

Очевидно, этого можно достигнуть, образуя направления по формуле

$$p_i = -P_q g_i \qquad \text{(IV, 102)}$$

где P_q — матрица ортогонального проектирования (III, 38)

$$P_q = I_n - N_q \left(N_q^{\mathrm{T}} N_q\right)^{-1} N_q^{\mathrm{T}} \qquad \text{(IV, 103)}$$

Направление p_i в данном случае является проекцией градиента функции на пересечение гиперповерхностей (IV,100) и дает направление наискорейшего убывания функции в многообразии L_q.

Ясно, что в точке минимума $\bar{x}$ должно выполняться соотношение:

$$P_q g(\bar{x}) = 0 \qquad \text{(IV, 104)}$$

Умножив выражение (IV, 103) на $g(\bar{x})$ и использовав соотношения (IV, 104), получим

$$g(\bar{x}) = N_q \gamma \qquad \text{(IV, 105)}$$

где $\gamma = \left(N_q^{\mathrm{T}} N_q\right)^{-1} N_q^{\mathrm{T}} g(\bar{x})$ — вектор, компонентами которого являются множители Лагранжа. Если вектор γ рассчитывается не в точке минимума, а в текущей точке x_k

$$\gamma = \left(N_q^{\mathrm{T}} N_q\right)^{-1} N_q^{\mathrm{T}} g_k \qquad \text{(IV, 106)}$$

то его компоненты называются оценками первого порядка множителей Лагранжа.

Формула (IV,102) лежит в основе метода проектирования градиента [88, с. 134—135]. Ясно, что по скорости сходимости этот метод эквивалентен методам градиента и наискорейшего спуска для случая отсутствия ограничений. Поэтому методы переменной метрики, дающие большую скорость сходимости, интересно распространить на данный случай.

Предположим вначале, что матрица A [см. выражение (III, 2)] известна и имеет ранг n. Поставим следующую задачу — пусть координаты точки x_i удовлетворяют уравнениям (IV, 100); необходимо найти точку минимума функции (III, 2) в многообразии L_q.

В точке минимума вектор $g(\bar{x})$ должен быть ортогонален L_q. Считая, что точка x_{i+1} соответствует точке минимума, подставим в соотношения (III, 8) $g_{i+1} = N_q a$, где a — некоторый вектор, координаты которого надо найти. Умножив обе стороны этого равенства на A^{-1}, получим:

$$x_{i+1} - x_i = A^{-1}(N_q a - g_i) \qquad \text{(IV, 107)}$$

Поскольку вектор $s_i = x_{i+1} - x_i \in L_q$, $N_q^{\mathrm{T}} s_i = 0$; отсюда, умножив равенство (IV, 107) на N_q^{T} слева, получим

$$N_q^{\mathrm{T}} A^{-1} N_q a - N_q^{\mathrm{T}} A^{-1} g_i = 0 \qquad \text{(IV, 108)}$$

Поскольку ранг N_q равен q, и ранг A^{-1} равен n, существует матрица, обратная к матрице $N_q^{\mathrm{T}} A^{-1} N_q$ [89], поэтому из уравнения (IV, 108) получим

$$a = \left(N_q^{\mathrm{T}} A^{-1} N_q\right)^{-1} N_q^{\mathrm{T}} A^{-1} g_i \qquad \text{(IV, 109)}$$

Компоненты вектора a называются оценками второго порядка множителей Лагранжа. Подставив это значение a в равенство (IV, 107), получим [89]:

$$x_{i+1} = x_i - \overline{P}_q A^{-1} g_i \qquad \text{(IV, 110)}$$

где $\overline{P}_q$ — обобщенный проекционный оператор

$$\overline{P}_q = I_n - A^{-1}N_q \left(N_q^{\tau}A^{-1}N_q\right)^{-1} N_q^{\tau} \qquad \text{(IV, 111)}$$

Если рассматривается случай минимизации неквадратичной функции и возможно вычисление гессиана G функции f, то на основе формулы (IV, 110) может быть построен аналог метода Ньютона для случая условной минимизации с линейными ограничениями. В этом случае итерационные формулы примут вид

$$x_{k+1} = x_k - \overline{P}_q G_k^{-1} g_k \qquad \text{(IV, 112)}$$

где

$$\overline{P}_q = I_n - G_k^{-1}N_q \left(N_q^{\tau}G_kN_q\right)^{-1} N_q^{\tau}$$

В случае отсутствия ограничений формула (IV,112) сводится к обычной формуле метода Ньютона. Рассмотрим теперь матрицу

$$\overline{\overline{P}}_q = I_n - HN_q \left(N_q^{\tau}HN_q\right)^{-1} N_q^{\tau} \qquad \text{(IV, 113)}$$

где H — произвольная матрица, ранг которой больше или равен q. Матрица $\overline{\overline{P}}_q$ получена формальной подстановкой в выражение (IV, 111) матрицы H вместо A^{-1}.

Покажем, что выполняются соотношения

$$n_i^{\tau}\overline{\overline{P}}_q = 0 \qquad i = 1, \ldots, q \qquad \text{(IV, 114)}$$

Действительно

$$N_q^{\tau}\overline{\overline{P}}_q = N_q^{\tau} - \left(N_q^{\tau}HN_q\right)\left(N_q^{\tau}HN_q\right)^{-1} N_q^{\tau} \equiv 0$$

Рассмотрим частный случай формулы (IV, 113), когда матрица N_q состоит из одного столбца n_q:

$$\overline{\overline{P}}_1 = I_n - \frac{Hn_q n_q^{\tau}}{n_q^{\tau}Hn_q} \qquad \text{(IV, 115)}$$

Введем в рассмотрение матрицу R

$$R = \overline{\overline{P}}_1 H = H - \frac{Hn_q n_q^{\tau}H}{n_q^{\tau}Hn_q} \qquad \text{(IV, 116)}$$

в которой матрица H удовлетворяет соотношениям

$$Hn_i = 0 \qquad i = \overline{1,\ q-1} \qquad \text{(IV, 117)}$$

Матрица R обладает свойством:

$$Rn_i = 0 \qquad i = \overline{1,\ q} \qquad \text{(IV, 118)}$$

Действительно, умножив выражение (IV,116) на n_i, получим

$$Rn_i = Hn_i - \frac{Hn_q n_q^{\tau}Hn_i}{n_q^{\tau}Hn_q} = 0$$

Для $i < q$ это равенство верно в соответствии с условиями (IV, 117), а для $i = q$ оно легко проверяется.

Рассмотрим теперь ряд методов минимизации функций многих переменных при наличии линейных ограничений.

Метод Гольдфарба [89]. Рассмотрим вначале алгоритм движения при неизменном базисе. Покажем, что последовательность точек, определяемая формулами (I, 39), (I, 41), будет лежать в многообразии L_q при соблюдении следующих условий:

1. Начальная точка $x_0 \in L_q$.
2. Матрица H_i в выражении (I, 41) определяется формулой (II, 90).
3. Матрица R в формуле (II, 90) удовлетворяет соотношениям (IV, 118), а коэффициенты c_{hi}, d_{hi} — соотношениям

$$c_{ki}^{\tau}N_q = d_{ki}^{\tau}N_q = 0 \qquad k = 1,\ 2 \qquad i = 0,\ 1, \ldots \qquad \text{(IV, 119)}$$

Доказательство проведем по индукции. Рассмотрим вначале первый шаг ($i = 0$). Из соотношений (IV,118) и (II,93) получим:

$$H_0 n_i = 0 \qquad K_0^m n_i = n_i \qquad m = 1,\ 2 \qquad i = \overline{1,\ q} \qquad \text{(IV, 120)}$$

отсюда, вектор $p_0 = -H_0^{\text{T}} g_0$ ортогонален векторам n_i, ($i = \overline{1, q}$). Следовательно, все точки полупрямой, начинающейся в точке x_0, ($x_0 \in L_q$) и проходящей вдоль вектора p_0, будут принадлежать L_q. В частности точка $x_1 \in L_q$.

Рассмотрим ($i + 1$)-й шаг. Пусть выполняется соотношение

$$H_i n_j = 0 \qquad K_i^m n_j = n_j \qquad m = 1,\ 2 \qquad j = \overline{1,\ q} \qquad \text{(IV, 121)}$$

и точка $x_{i+1} \in L_q$. Покажем, что $x_{i+2} \in L_q$. Умножив справа равенства (II, 90), (II, 91) на n_j и использовав выражения (IV, 119), (IV, 121), легко получить равенства

$$H_{i+1} n_j = 0 \qquad K_{i+1}^m n_j = n_j \qquad m = 1,\ 2 \qquad j = \overline{1,\ q} \quad \text{(IV, 122)}$$

Отсюда, вектор $p_{i+1} = -H_{i+1}^{\text{T}} q_{i+1}$ будет ортогонален векторам n_j, ($j = \overline{1, q}$). Следовательно, все точки полупрямой, начинающейся в точке x_{i+1} и проходящей вдоль вектора p_{i+1}, будут принадлежать L_q. В частности $x_{i+2} \in L_q$. Таким образом, все точки x_i, ($i = 0, 1, \ldots$) лежат в L_q. Матрица R может быть определена с помощью формулы (IV, 116).

Теперь покажем, что для любой формулы семейства Хуанга (III, 95) условия (IV, 119) будут выполняться. Действительно, в этом случае коэффициенты c_{ki}, d_{ki} определяются по формулам (III, 100), (III, 101), и соотношения (IV, 119) примут вид

$$\alpha_{ji} y_i^{\text{T}} H_i N_q + \beta_{ji} s_i^{\text{T}} N_q = 0, \quad \gamma_{ji} y_i^{\text{T}} H_i N_q + \delta_{ji} s_i^{\text{T}} N_q = 0 \qquad \text{(IV, 123)}$$

В точке x_0 эти соотношения являются следствием условия (IV, 118). Пусть в i-той точке они выполняются: покажем, что они будут выполняться в ($i + 1$)-й точке. В i-той точке выполняются условия (IV, 121), поэтому $H_i N_q = 0$, а $s_i^{\text{T}} N_q = 0$ (направление движения в i-той точке лежало в L_q), что и доказывает тождество (IV, 123).

Направления p_i строятся с помощью формулы (I, 41), в которой H_i удовлетворяют уравнению (II,32), следовательно, эти направления будут сопряженными. С другой стороны, как мы показали, все векторы p_i будут лежать в многообразии L_q размерности $n - q$. Отсюда (см. с. 82), минимум в многообразии L_q будет найден за число шагов, не большее, чем $n - q$. В дальнейшем матрицы H_i, которые обеспечивают движение в L_q, будем обозначать через H_i^q. Нижний индекс здесь по-прежнему обозначает номер итерации, а верхний — число линейных ограничений, образующих линейное многообразие L_q.

Остановимся теперь на алгоритме выбора шага вдоль направления p_i. Очевидно, шаг вдоль направления p_i должен быть меньше или равен расстоянию от точки x_i до ближайшего неактивного ограничения. Легко видеть, что точка пересечения луча, исходящего из точки x_i по направлению p_i [см. формулы (I, 41)], с некоторой гиперплоскостью

$$n_q^{\text{T}} x - b_q = 0 \qquad q \geqslant m + 1 \qquad \text{(IV, 124)}$$

определяется значением коэффициента $\alpha_i = \alpha_{iq}$, где

$$\alpha_{iq} = \frac{b_q - n_q^{\text{T}} x_i}{n_q^{\text{T}} p_i} \qquad \text{(IV, 125)}$$

Найдем величину

$$\bar{\alpha}_i = \min_q \alpha_{iq} \qquad \text{(IV, 126)}$$

Отсюда коэффициент α_i в выражении (I, 39) надо выбирать в соответствии с формулой

$$\alpha_i = \begin{cases} \alpha_i^*, & \text{если } \bar{\alpha}_i > \alpha_i^* \qquad \text{(IV, 127)} \\ \bar{\alpha}_i, & \text{если } \bar{\alpha}_i < \alpha_i^* \qquad \text{(IV, 128)} \end{cases}$$

где α_i^* — значение α_i, соответствующее оптимальной точке на направлении p_i. Ясно, что если выполняется условие (IV, 127), то базис ограничений остается без изменений, если же выполняется условие (IV, 128), то в базис вносится ограничение с номером j, для которого $\alpha_{ij} = \bar{\alpha}_i$, т. е. уравнение той гиперплоскости, на которую попадет изображающая точка в результате итерационного шага (I, 39). Гиперплоскость с номером l удаляется из базиса, если в i-той точке выполняется неравенство

$$\left\| H_i^q g_i \right\| \leqslant \frac{1}{2} \gamma_l b_{ll}^{-1/2} \qquad \text{(IV, 129)}$$

где

$$\gamma_l b_{ll}^{-1/2} = \max_i \left(\gamma_i b_{ii}^{-1/2} \right), \quad i = m+1, \ldots, q$$

Здесь в b_{ii} — i-тый диагональный элемент матрицы $\left(N_q^{\mathrm{T}} N_q\right)^{-1}$; γ_i — множители Лагранжа [см. выражение (IV, 106)].

Рассмотрим алгоритмы изменения матрицы H при увеличении и уменьшении числа активных ограничений.

Предположим, что базис одновременно может увеличиться только на одно активное ограничение (маловероятно, что изображающая точка во время поиска может сразу попасть на пересечение двух гиперплоскостей, соответствующих активным ограничениям). Итак, пусть до некоторой точки x_k базис состоял из $q - 1$ линейных ограничений, векторы нормалей к которым образовали матрицу N_{q-1}, а в этой точке к базису должно быть добавлено уравнение гиперплоскости (IV, 124). Пусть до точки x_k было проведено $k - 1$ итераций. Точка x_k лежит на прямой, проходящей вдоль вектора $p_{k-1} = -H_{k-1}^{q-1} g_{k-1}$. Как было показано, чтобы поисковые точки, получаемые с помощью формул (I, 39), (I, 41), в которых H_i вычисляются по формуле (II, 107), лежали в многообразии L_q, необходимо, чтобы матрица R удовлетворяла условиям (IV, 118). Образуем матрицу R в соответствии с формулой (IV, 116), в которой H взята равной H_{k-1}^{q-1}. Обозначив матрицу R через H_k^q, получим

$$H_k^q = R = H_{k-1}^{q-1} - \frac{H_{k-1}^{q-1} n_q n_q^{\mathrm{T}} H_{k-1}^{q-1}}{n_q^{\mathrm{T}} H_{k-1}^{q-1} n_q} \qquad \text{(IV, 130)}$$

Поскольку матрица H_{k-1}^{q-1} удовлетворяет соотношениям (IV, 117), матрица H_k^q будет удовлетворять соотношениям (IV, 118). Пусть на i-той итерации в соответствии с критерием (IV, 129) ограничение, имеющее номер q, должно стать неактивным, следовательно, из базиса следует удалить уравнение q-ой гиперплоскости. Матрицу H_i^q предлагается изменять по формуле:

$$H_i^{q-1} = H_i^q + \frac{P_{q-1} n_q n_q^{\mathrm{T}} P_{q-1}}{n_q^{\mathrm{T}} P_{q-1} n_q} \qquad \text{(IV, 131)}$$

где P_{q-1} — проекционная матрица, соответствующая оставшимся активными ограничениям. Если $f(x)$ является квадратичной функцией, рассмотренный алгоритм обеспечит решение задачи (IV, 97)—(IV, 99) за конечное число итераций. Число итераций будет больше n, как это было в случае безусловной минимизации. Для применения этого алгоритма в случае минимизации неквадратичных функций его следует дополнить некоторыми операциями, близкими к тем, которые применялись в аналогичном случае в задаче безусловной минимизации (например, процедурой «обновления» матрицы H_i^q в некоторых ситуациях, и др.).

Отметим, что для преобразования матриц H_i^q может быть использована любая из формул семейства Хуанга (III, 95), а не только преобразование DFP (см. работу Гольдфарба [89]). Остановимся теперь на недостатках метода Гольдфарба. Пусть до точки x_k (см. рис. 25, точка B) поиск шел во всем пространстве, а начиная с этой точки он должен пойти в многообразии L_1. Матрица H_k^0, полученная в этой точке, удовлетворяет уравнению (II, 32). Ясно, что матрица H_k^1, полученная с помощью выражения (IV,130), не будет решением системы (II, 32), следовательно, информация,

содержавшаяся в H_k^0 относительно гессиана функции $f(x)$ в результате преобразования (IV,130) будет искажена. Аналогичным образом, преобразование (IV, 131) будет искажать информацию о гессиане функции $f(x)$.

Метод Муртага—Саджента. Отмеченные недостатки метода Гольдфарба связаны с тем, что в нем матрица H_i^k фактически аппроксимирует произведение матриц $\overline{P}_q G_k^{-1}$ [см. выражение (IV,112)]. Пытаясь избавиться от этого недостатка, Муртаг и Саджент предложили метод [90, с. 88], в котором матрицы $\overline{P}_q$ и G_k^{-1} аппроксимируются отдельно. Этот метод может быть получен, если в формулах (IV, 107), (IV, 109) заменить A^{-1} ее аппроксимацией

$$x_{i+1} = x_i + H_i (N_q a - g_i) \qquad a = \left(N_q^{\mathrm{T}} H_i N_q\right)^{-1} N_q^{\mathrm{T}} H_i g_i$$

В данном методе матрицы H_i, $\left(N_q^{\mathrm{T}} H_i N_q\right)^{-1}$ пересчитываются отдельно. Матрица H_i может пересчитываться с помощью любой из формул семейства (II, 90), (II, 91), или (III, 95). Отсюда, при изменении базиса информация о гессиане, содержащаяся в H_k, не будет искажаться. Однако, метод Муртага — Саджента имеет один существенный недостаток, который присущ и методу Гольдфарба. Если ограничение (IV, 124) было активно на достаточно большом числе итераций, то в матрице H_k, аппроксимирующей G_k, будет отсутствовать информация о кривизне $f(x)$ по направлению нормали к гиперплоскости (IV, 124). Поэтому в точке x_l схода с ограничения (IV, 124) H_l будет плохо аппроксимировать G_l^{-1}.

Квазиньютоновский метод 1-го рода. * Вначале рассмотрим алгоритм преобразования матрицы H_j, обеспечивающий движение в заданном многообразии. Для определенности будем предполагать, что в $(j+1)$-ой точке активными должны быть первые $(q-1)$ ограничения (IV, 100), т. е. что в точке x_{j+1} должны выполняться соотношения (IV, 101). Матрицу H_{j+1} будем определять из условия, чтобы она удовлетворяла квазиньютоновскому условию 1-го рода (II,31) в $(j+1)$-ой точке, была симметрична и обеспечивала выполнение соотношений (IV, 101). С использованием соотношения (II, 56) задача поиска матрицы H_{j+1} сформулируется как задача (III, 64), дополненная соотношениями

$$g_{j+1}^{\mathrm{T}} D n_i = b_i \qquad b_i = -g_{j+1}^{\mathrm{T}} H_j n_i \qquad i = \overline{1,\ q-1} \qquad \text{(IV, 132)}$$

которые получаются подстановкой в соотношения (IV, 101) формулы для p_k, полученной из (I, 41) и для H_{j+1} из (II, 56). Проведя вывод, совершенно аналогичный выводу формул (III, 63), (III, 65), легко получить выражение для D:

$$D = D_{Gr} + \sum_{k=1}^{q-1} \mu_k \left(\frac{\gamma_k M_j}{(y_j^{\mathrm{T}} y_j)^2} - \frac{D_k M_j}{y_j^{\mathrm{T}} y_j} - \frac{M_j D_k}{y_j^{\mathrm{T}} y_j} + D_k \right) \qquad \text{(IV, 133)}$$

где матрица D_{Gr} определяется с помощью формулы (III, 65); $M_j = y_j y_j^{\mathrm{T}}$;

$$D_k = g_{j+1} n_k^{\mathrm{T}} + n_k g_{j+1}^{\mathrm{T}} \qquad \gamma_k = y_j^{\mathrm{T}} \left(g_{j+1} n_k^{\mathrm{T}} + n_k g_{j+1}^{\mathrm{T}}\right) y_j \qquad \text{(IV, 134)}$$

Множители Лагранжа μ_k, соответствующие условиям (IV, 132), находятся решением системы линейных уравнений:

$$g_{j+1}^{\mathrm{T}} D_{Gr} n_i + \sum_{k=1}^{q-1} \mu_k \left[\frac{2\gamma_k}{(y_j^{\mathrm{T}} y_j)^2} g_{j+1}^{\mathrm{T}} M_j n_i - \frac{1}{y_j^{\mathrm{T}} y_j} g_{j+1}^{\mathrm{T}} D_k M_j n_i - \right.$$
$$\left. - \frac{1}{y_j^{\mathrm{T}} y_j} g_{j+1}^{\mathrm{T}} M_j D_k n_i + g_{j+1} D_k n_i = b_i \right. \qquad \text{(IV, 135)}$$

Рассмотрим общий алгоритм решения задачи (IV, 97), (IV, 99). Пусть в j-той точке базис состоял из $(q-1)$-го линейного ограничения, а в $(j+1)$-ой точке в ба-

* Раздел написан совместно с Е. М. Михайловой.

зис добавилось одно ограничение, номер которого q. Это значит, что условия (IV, 132) будут дополнены условиями $g_{j+1}^{T}Dn_q = b_q$, и тогда в выражение (IV, 133) для D добавится один член, соответствующий этому условию. Поскольку матрица D выбирается из условия, чтобы ее норма Фробениуса была минимальна, то можно ожидать, что информация о гессиане функции $f(x)$, содержащаяся в матрице H_j, в максимальной степени сохранится в матрице H_{j+1}. Аналогичная ситуация будет и в точке x_l, где линейное ограничение (IV, 124) должно быть выведено из базиса. В этой точке условие $g_l^{T}Dn_q = b_q$ должно быть удалено из соотношений (IV, 132). В этом случае также и в матрице H_l должна в максимальной степени сохраниться информация о гессиане, которая содержалась в матрице H_{l-1}. Конечно, если ограничение (IV,124) было активно на достаточно большом числе итераций, то в матрице H_{l-1} (так же, как и в случае применения методов Гольдфарба и Муртага — Сарджента) будет отсутствовать информация о кривизне $f(x)$ вдоль нормали к гиперплоскости (IV,124). Однако в этом случае можно поступить следующим образом. Вычислим значение градиента функции $f(x)$ в некоторой точке, находящейся на нормали n_q и после этого подсчитаем величину v_k

$$v_k = g(x_k + hn_q) - g(x_k)$$

где h некоторый скаляр. Рассмотрим теперь соотношение

$$H_k v_k = hn_q \qquad (IV,\ 136)$$

Ясно, что оно содержит информацию о кривизне функции $f(x)$ вдоль нормали n_q. Поэтому, поскольку квазиньютоновское условие (II, 31) содержит мало информации о кривизне функции $f(x)$ вдоль нормали к гиперплоскости (IV, 124), то наряду с соотношениями (IV, 132) добавим к задаче (III, 64) условие (IV, 136). Решив полученную задачу, мы найдем вид матрицы H_k, которая будет содержать информацию о кривизне $f(x)$ вдоль нормали n_q. Следует, конечно, отметить, что соотношение (IV, 136) получается ценой одного дополнительного вычисления градиента.

У этого подхода имеется еще один недостаток. В его основе лежит формула Гринштадта, которая, как показывает опыт решения задач безусловной минимизации, не дает хороших результатов [27]. Интересно обобщить этот подход на более эффективные методы, типа DFP, BFGS. Для этого так же как и при выводе квазиньютоновских методов безусловной минимизации 1-го рода необходимо заменить критерий (III, 64) нормой Фробениуса взвешенной суммы (III, 76). В этом случае, применяя подход, использованный при решении задачи (III, 68), можно получить выражение для H_{j+1}.

У рассмотренных подходов имеется один недостаток — на каждом шаге приходится решать систему линейных уравнений (IV, 135), поэтому может оказаться целесообразным использовать их только в случае, когда число активных ограничений в каждой точке будет мало, а следовательно, будет мала размерность системы линейных уравнений (IV,135). Правда, указанный недостаток имеет и положительную сторону. Действительно, решение системы (IV, 135) на каждом шаге гарантирует ортогональность векторов p_i нормалям к гиперплоскостям, входящим в базис, что не дает возможности накапливаться ошибкам округления и приводить к нарушению этой ортогональности, как это может происходить при применении методов Гольдфарба и Муртага — Сарджента.

Метод обобщенного приведенного градиента

Рассмотрим задачу нелинейного программирования (IV, 1)—(IV, 4), в которой ограничения (IV,4) имеют вид

$$a_i \leqslant x_i \leqslant b_i \qquad i = \overline{1,\ n} \qquad (IV,\ 137)$$

Основная идея метода обобщенного приведенного градиента (МОПГ) [92, с. 49] состоит в разделении совокупности переменных x_i, $(i = \overline{1,\ n})$ на группу зависимых и группу независимых переменных. Число зависимых переменных равно p — числу ограничений (IV, 3). Для определенности в качестве зависимых переменных выберем

переменные $x_1, \ldots, x_p$, в качестве независимых переменных $x_{p+1}, \ldots, x_n$. Введем обозначения:

$$y_i = x_i \qquad (i = \overline{1,\ p}) \qquad u_i = x_{p+i} \qquad (i = 1, \ \ldots, \ n-p) \quad (\text{IV},\ 138)$$

Обозначим через y, u векторы с компонентами $y_1, \ldots, y_p$, $u_1, \ldots, u_{n-p}$ соответственно. Тогда задача (IV, 1)—(IV, 4) запишется в виде

$$\min_{y,\ u} f(y,\ u) \qquad (\text{IV},\ 139)$$

$$\varphi_i(y,\ u) = 0 \qquad i = \overline{1,\ p} \qquad (\text{IV},\ 140)$$

$$a_i \leqslant u_i \leqslant b_i \qquad i = \overline{1, n-p} \qquad (\text{IV},\ 141)$$

$$a_i \leqslant y_i \leqslant b_i \qquad i = \overline{1,\ p} \qquad (\text{IV},\ 142)$$

При фиксированном u соотношения (IV, 140) могут рассматриваться как система p нелинейных уравнений с p неизвестными компонентами вектора y. Будем исходить из предположения, что матрица Якоби системы (IV, 140) по переменным y не равна нулю. Отсюда равенства (IV,140) определяют переменные y как неявные функции переменных u: $y = y(u)$. Будем теперь искать минимум f как сложной функции переменных u

$$f^*(u) = f[y(u),\ u] \qquad (\text{IV},\ 143)$$

Для простоты рассмотрим случай, когда ограничения (IV,142) отсутствуют. Тогда задача (IV, 139)—(I, 141) сводится к следующей:

$$\min_{u} f[y(u),\ u] \qquad (\text{IV},\ 144)$$

$$a \leqslant u \leqslant b \qquad (\text{IV},\ 145)$$

Отсюда уже видно основное преимущество подхода, основанного на разделении первоначальных переменных на зависимые и независимые; он позволяет во-первых, исключить из задачи минимизации ограничения типа равенств; во-вторых, уменьшить число поисковых переменных. Правда, при этом возникает новая задача определения величин y при фиксированном u. Для решения задачи (IV, 144), (IV, 145) необходимо иметь:

1. Алгоритм определения переменных y как решение системы (IV, 140) при фиксированных u.

2. Алгоритм вычисления производных сложной фукции (IV, 143) по иезависимым переменным.

3. Алгоритм безусловной минимизации, модифицироваиный для учета ограиичений (IV, 145).

Для определения y можно использовать прием лииеаризации [92, с. 49]. Применяя правила диффереицирования сложных и неявных фуикций, легко получить формулы для определения производных функции (IV, 143) по перемениым u [92, с. 49]. Для решения задачи (IV, 144), (IV, 145) используется метод сопряженных градиентов, модифицированный для учета ограиичеиий (IV,145) (МОПГ); ои был предложеи в 1968 г. и является обобщением метода приведенного градиента, разработанного Вольфом [93] для решения задачи (IV, 1), (IV, 3), (IV, 141) с линейиыми ограничениями (IV, 3), на случай нелинейных ограничений (IV, 3). Вместе с тем следует отметить, что при решении задач оптимизации в химической технологии этот подход введения зависимых и независимых переменных для исключения ограничений типа равенства фактически использовался уже в начале 60-х годов. Причем в качестве зависимых переменных обычно выбирались переменные состояния, в качестве независимых — управления [94], а в качестве ограничений типа равенств выступали математические модели блоков и уравнения связи. На основе этого подхода был дан способ вычисления градиента функции (IV,143) для ряда типовых схем [95, 96]. Имеется также более удобный способ вычисления производных функций (IV, 143) для общего случая [97]. В чистом виде МОПГ эквивалентен задаче 2 оптимизации ХТС [см. соотношение (I, 71), (I, 72)], либо задаче 1 [см. соотношения (I, 64)—(I, 66)], когда ограничения (I, 10) отсутствуют.

Оптимизация процесса полимеризации изопрена в производстве синтетического каучука *

Технологическая схема процесса полимеризации изопрена в производстве синтетического каучука состоит из 6—8 параллельно работающих батарей, каждая из которых является каскадом из двух непрерывно действующих реакторов с мешалкой (рис. 26). В реакторах идет процесс полимеризации изопрена в растворе изопентана [100, с. 236].

Шихта G (раствор изопрена в изопентане), поступающая в отделение полимеризации при температуре T_0, концентрации изопрена m_0 и концентрации водорода H_0, распределяется по работающим батареям. В первые реакторы батарей подается катализатор $G_к$, в рубашку реакторов — хладоагент $G_{хл}$. Выходные потоки всех батарей смешиваются и поступают в отделение выделения и сушки. Непрореагировавший изопрен (мономер в возвратной фракции $m_{в.ф}$) из отделения ректификации вновь поступает на вход батарей. Изменение во времени характеристик реакторов процесса, а также изменение количества примесей требуют оптимизации стационарного режима действующего процесса полимеризации, проводимой через определенные промежутки времени.

В качестве управляющих выбраны следующие 15 переменных процесса (число работающих батарей $N_б = 6$): концентрация изопрена в шихте m_0, температура шихты T_0, концентрация водорода в шихте H_0, нагрузка на батарею G^j, расход катализатора на батарею $G_к^j$ $(j = \overline{1, N_б})$.

Математическая модель процесса состоит из совокупности математических моделей отдельных аппаратов (реакторов полимеризации, смесителя и делителя потока) и уравнений связи между ними. Уравнения связи здесь очень просты, так что выписывать их не будем.

В делителе поступающая шихта G разделяется на $N_б$ параллельно работающих батарей:

$$G = \sum_{j=1}^{N_б} G^j$$

где G^j — поток на входе j-ой батареи. Естественно, что концентрация изопрена в шихте m_0 и температура шихты T_0 в потоках после делителя равны соответственно концентрации изопрена и температуре входного потока.

Математическая модель реактора состоит из уравнений тепло- и массопередачи, а также зависимостей вязкости (по Муни) полимера от режимных параметров процесса полимеризации. В дополнение к известной модели процесса [99, с. 16] введены материальный баланс по водороду, уравнения смешения для мономера в возвратной фракции $m_{в.ф}$ и показателя качества Муни $M_{г.к}$ готового каучука. При записи модели сразу учтем, что выходные переменные i-го реактора являются входами в $(i + 1)$-й реактор.

* Раздел написан совместно с С. Л. Подвальным и Е. М. Михайловой.

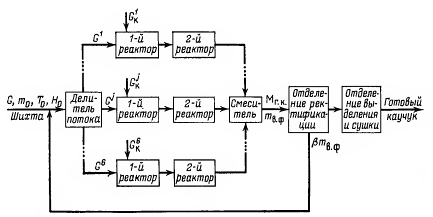


Рис. 26. Схема процесса полимеризации изопрена.

В статике математическая модель реактора полимеризации (в безразмерном виде) записывается следующим образом:

$$\begin{cases} 1{,}6\,\dfrac{G^j}{V}\left(m_{i-1}^j - m_i^j\right) - R_i^j f\left(m_i^j\right) = 0 \\ 1{,}6\,\dfrac{G^j}{V}\left(T_{i-1}^j - T_i^j\right) + a_1 R_i^j f\left(m_i^j\right) + k_{\text{хл}}^j\left(T_{\text{хл}} - T_i^j\right) = 0 \\ 10^{-3}\,\dfrac{G_k^j}{V}\,n_{\text{ок. р}} - 1{,}6\,\dfrac{G^j}{V}\,n_i^j = 0, \; i = 1 \\ 1{,}6\,\dfrac{G^j}{V}\left(n_{i-1}^j - n_i^j\right) = 0, \; i \geqslant 2 \\ 1{,}6\,\dfrac{G^j}{V}\left(H_{i-1}^j - H_i^j\right) - 1{,}2 H_i^j = 0 \\ M_i^j x_i^j - M_{i-1}^j x_{i-1}^j - \left(x_i^j - x_{i-1}^j\right) M_i^{*j} = 0 \end{cases} \qquad \text{(IV, 146)}$$

$$R_i^j = \frac{b^j}{t_{0,5}}\left(9{,}15\cdot 10^{-2} T_i^j + 1{,}5\right)\left(5{,}8 - 0{,}185 m_0\right) 14{,}42\cdot 10^{-3} n_i^j$$

$$f\left(m_i^j\right) = \left[1 - 0{,}5\left(m_i^j/m_0\right)^2\right] m_i^j$$

$$k_{\text{хл}}^j = 1{,}8\,\frac{G_{\text{хл}}^j}{V}\left[1 - \exp\left(-0{,}0342\cdot 40/G_{\text{хл}}^j\right)\right]$$

$$x_i^j = \left(m_0 - m_i^j\right)/m_0 \qquad D_{\text{к}}^j = G_{\text{к}}^j n_{0\,\text{к. р}}\cdot 10^{-2}/\left(G^j m_0\right)$$

$$M_i^{*j} = a_2 - 0{,}25 x_i^j - 1{,}19 T_i^j + (m_0 - 12) + \lambda\left(V/G^j\right) + \exp\left(-0{,}45 m_0 H_i^j\right)$$

где m_i^j — концентрация изопрена; T_i^j — температура; n_i^j — концентрация катализатора; H_i^j — концентрация водорода; M_i^j — вязкость по Муни; $T_{\text{хл}}$ — температура хладагента; $G_{\text{хл}}^j$ — расход хладагента; V — объем реактора; $n_{\text{ок. р}}$ — концентрация катализатора в каталитическом растворе; $t_{0,5}$ — период полупревращения; D_k^j — до-

зировка катализатора. Нижний индекс i ($i = 1, 2$) у всех переменных обозначает номер реактора, а верхний j ($j = \overline{1, N_б}$) — номер батареи. Поскольку модель не отражает влияния примесей на выходные переменные реактора, то требуется подстройка модели. Настроечными параметрами при этом являются a_1, a_2, b^j ($j = \overline{1, N_б}$).

Уравнения математической модели реактора при заданных значениях входных и управляющих переменных представляют собой систему из 5-ти уравнений с 5-ю неизвестными. Решение этой системы проводилось методом Ньютона. Благодаря хорошему начальному приближению на расчет реактора в среднем требовалось 2—3 итерации.

Математическая модель смесителя имеет вид:

$$m_{в.\,ф.} = \frac{\sum_{j=1}^{N_б} G^j m_2^j}{\sum_{j=1}^{N_б} G^j \left[1 - \left(m_0 - m_2^j\right) \cdot 10^{-2}\right]} \qquad M_{г.\,к.} = \frac{\sum_{j=1}^{N_б} G^j x_2^j M_2^j}{\sum_{j=1}^{N_б} G^j x_2^j}$$

где $m_{в.\,ф}$ — концентрация мономера в возвратной фракции; $M_{г.\,к}$ — вязкость по Муни готового каучука. Поскольку получаемый каучук должен обладать определенными свойствами, среди которых одним из основных является $M_{г.\,к}$, в качестве критерия оптимизации может быть выбрано квадратичное отклонение от заданного качества:

$$\Delta M_{г.\,к} = \left(M_{г.\,к} - M_{г.\,к}^{зад}\right)^2 / \left(M_{г.\,к}^{зад}\right)^2 \qquad (IV,\ 147)$$

где $M_{г.\,к}^{зад}$ — заданное значение вязкости по Муни при соответствующем ограничении на производительность батарей P. В том случае, когда вязкость по Муни задана не жестко, а в каком-то интервале, задача оптимизации решается по критерию максимума производительности работающих батарей:

$$P = \sum_{j=1}^{N_б} G^j \left(m_0 - m_2^j\right) \cdot 10^{-2} \qquad (IV,\ 148)$$

В задаче оптимизации должны выполняться ограничения типа (I, 9) на следующие входные и выходные переменные схемы m_0, T_0, H_0, G^j, G_k^j, T_i^j, D_k^j. Кроме того, должны выполняться следующие ограничения

$$\sum_{j=1}^{N_б} G^j = G \qquad T_1^1 = T_1^l \qquad l = \overline{2,\ N_б} \qquad (IV,\ 149)$$

При использовании критерия (IV,147), надо учитывать ограничение:

$$m_{в.\,ф}^{*} \leqslant m_{в.\,ф} \leqslant m_{в.\,ф}^{**} \qquad (IV,\ 150)$$

Причем, можно показать, что верхнее ограничение приблизительно эквивалентно ограничению на производительность, а нижнее обеспечивает недопущение формирования застойной зоны в реакторе и зарастания трубопроводов. При использовании критерия (IV, 148)

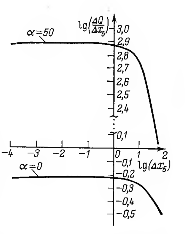


Рис. 27. Зависимость разностной аппроксимации производной от величины приращений аргумента.

в соотношении (IV, 150) остается только нижнее ограничение и добавляется ограничение по Муни:

$$M^{*}_{\text{г. к}} \leqslant M_{\text{г. к}} \leqslant M^{**}_{\text{г. к}} \qquad \text{(IV, 151)}$$

Итак, задача оптимизации стационарных режимов процесса полимеризации формулируется следующим образом: требуется определить такие значения управляющих переменных процесса (m_0, T_0, H_0, G^j, G^j_k, $j = \overline{1, N_б}$), которые минимизируют (максимизируют) один из критериев (IV, 147), (IV, 148) при соблюдении упомянутых ограничений на входные и выходные переменные. Вследствие малости интервала, задаваемого ограничением (IV, 150), величина $m_{\text{в. ф}}$ меняется незначительно и в первом приближении (при оптимизации схемы) рецикл можно не учитывать. Поэтому в дальнейшем изложении (за исключением случая, который будет оговорен особо) при оптимизации схемы не будем учитывать рецикл.

Целевую функцию будем обозначать через $Q(x, y)$, здесь x — вектор управляющих переменных, его размерность $N = 3 + 2N_б$; y — вектор фазовых переменных, его размерность $M = 10N_б$. Производные целевой функции вычисляются двумя способами. По первому способу производные вычисляются приближенно с помощью разностей (I, 49). Приращения Δx выбираются из области значений Δx, где величина $\Delta Q/\Delta x$ практически постоянна (рис. 27). Последовательность для Δx строится следующим образом:

$$\Delta x_j^{(l)} = \Delta x_j^{(l-1)}/10 \qquad l = 1,\ 2,\ \ldots;\ j = \overline{1, N} \qquad \text{(IV, 152)}$$

Выбираем Δx в середине участка его изменения, где правая часть выражения (I, 49) принимает приблизительно постоянное значение. Так, на рис. 27 кривые ($\alpha = 50$; 0) зависимости логарифма выражения (I, 49) от логарифма Δx_5 на участке $-3{,}5 \leqslant \lg(\Delta x_5) \leqslant -0{,}5$ принимают почти постоянные значения, и Δx_5 выбрано равным $10^{-3}x_5$; α — коэффициент штрафа. Выбор больших значений Δx приводит к грубой оценке производных вследствие погрешности метода конечных разностей. При малых Δx могут искажаться значения производных (вплоть до перемены знака) поскольку расчет схемы осуществляется с некоторой погрешностью. По второму способу производные вычисляются с помощью сопряженного процесса [3, с. 142]. Время, затрачиваемое на вычисление производных с помощью сопряженного процесса, составляет $1{,}6t_м$ (где $t_м$ — время расчета математической модели), тогда как при вычислении производных с помощью разностей эта величина составляет $(N + 1)\,t_м$.

Таблица 26. **Результаты оптимизации процесса полимеризации изопрена при двух способах вычисления производных целевой функции (IV,181)**

Метод безусловной оптимизации	Результаты оптимизации, полученные при использовании			
	конечно-разностного вычисления производных		метода сопряженного процесса	
	K_f	K_p	K_f	K_p
Гаусса—Зейделя	181	7	—	—
наискорейшего спуска	100	4	25	4
DFP	80	3	20	4

Таблица 27. **Результаты оптимизации стационарных режимов процесса полимеризации изопрена; критерий оптимизации [см. формулу (IV,181)]**

Метод безусловной оптимизации	$\frac{R_{opt}}{R_0}$	K_f	K_p	r
DFP	5,281	185	21	6
Ньютона	5,280	173	18	5
Гаусса—Зейделя	5,280	849	32	5
наискорейшего спуска	5,280	601	69	8
сопряженных градиентов	5,280	209	19	5

В табл. 26 приведены результаты сравнения двух способов вычисления производных целевой функции [критерий (IV, 147)]. Использовались следующие три метода безусловной оптимизации: безградиентный Гаусса—Зейделя, наискорейшего спуска и DFP. Применение метода сопряженного процесса позволяет сократить число вычислений целевой функции приблизительно в четыре раза. Для учета ограничений использовался метод штрафов, при котором проводилась безусловная минимизация функции (IV, 47) для некоторой последовательности значений параметров α_k^r, β_k^r, где r — номер итерации метода штрафов ($r = 0, 1, 2, \ldots$); $\alpha_k^{r+1} = q\alpha_k^r$; $\beta_k^{r+1} = q\beta_k^r$. Для ускорения сходимости метода штрафов применялся метод квадратичной экстраполяции [98, с. 361].

Численная процедура решения задачи оптимизации является четырехуровневой, причем 1-й уровень представляет собой расчет реактора; 2-й уровень — расчет схемы при фиксированных управляющих переменных: 3-й уровень — безусловную минимизацию, 4-й уровень — изменение штрафных коэффициентов.

В табл. 27 приведены результаты оптимизации стационарных режимов процесса полимеризации изопрена при следующих значениях параметров: $\alpha_k^{(0)} = 10$; $q = 5$ и приняты следующие обозначения: R_0, R_{opt} — соответственно начальное и оптимальное значения целевой функции; r — число итераций метода штрафов; производные вычислялись методом сопряженного процесса.

Кроме того, проводилась оптимизация схемы с рециклом. Для учета рецикла использовались три следующих способа, проиллюстрированные данными табл. 28: 1 — расчет рецикла с использованием метода простой итерации; 2 — расчет рецикла методом простой итерации с использованием приема прогнозирования (II, 200); 3 —

использование подхода, сводящего задачу оптимизации замкнутой ХТС к последовательности задач оптимизации разомкнутой ХТС следующим образом. Соотношение (II, 200) в линейном приближении дает решение системы уравнений замкнутой ХТС. В связи с этим проведем оптимизацию разомкнутой ХТС, предположим, что концентрация изопрена в рецикле вычисляется как явная функция вектора поисковых переменных u_{i+1} с помощью соотношения (II, 200). После того, как будет найден оптимум, необходимо при оптимальном векторе поисковых переменных решить систему уравнений замкнутой ХТС и в новой точке заново провести описанную процедуру и т. д. Решение задачи проводится по критерию (IV, 148). Из таблицы видно, что использование приема прогнозирования позволило уменьшить время, затрачиваемое на решение задачи оптимизации, в два раза, а использование приема линеаризации — в три раза.

Таблица 28. **Результаты оптимизации стационарного режима процесса полимеризации (с учетом рецикла)**

Номер способа учета рецикла	K_f	K_p	r	Машинное время, с
1	227	36	6	187,1
2	216	33	5	97,1
3	168	28	6	56,3

*Оптимизация конструкционных параметров теплообменной системы**

Рассматривается задача оптимизации теплообменной системы (ТС), показанной на рис. 28 и являющейся частью схемы некоторого производства [102]. ТС состоит из двенадцати теплообменников, двух делителей потоков — Д и смесителя С, фиктивных блоков ФБ, отражающих изменение температуры и давления в других аппаратах системы. Аппараты Т-2, Т-7, Т-8, Т-11, Т-12 осуществляют теплообмен между газом и водой, аппараты Т-3 и Т-4 выполнены в виде коробов с пакетами петлеобразных труб внутри, а остальные аппараты — обычные кожухотрубные теплообменники. Предполагаются заданными температуры потоков T_i^* на выходе ТС, а также общий допустимый перепад давления на линиях технологических газов Δp^* (I), газов среднего давления Δp^* (II) и газов низкого давления Δp^* (III). Для математического описания теплообменных процессов был использован метод [103], позволяющий учесть отклонения схемы взаимного движения теплоносителей от удельного прямотока или противотока. Соответствующие уравнения имеют вид:

$$T_{\text{хол}} = T_{\text{хол, 0}} + (T_{\text{гор, 0}} - T_{\text{хол, 0}}) \frac{1 - \exp\{-D[1 - \omega(1 - 2\varphi)]\}}{1 + \omega(1 - \varphi) - \omega\varphi E} \quad \text{(IV, 153)}$$

$$T_{\text{гор}} = T_{\text{гор, 0}} - (T_{\text{гор, 0}} - T_{\text{хол, 0}})\,\omega \quad \text{(IV, 154)}$$

где T и T_0 — температуры на выходе и входе соответственно, К; φ — коэффициент, определяющий зависимость характера тепло-

* Раздел написан совместно с Г. В. Михайловым и В. С. Витковым.

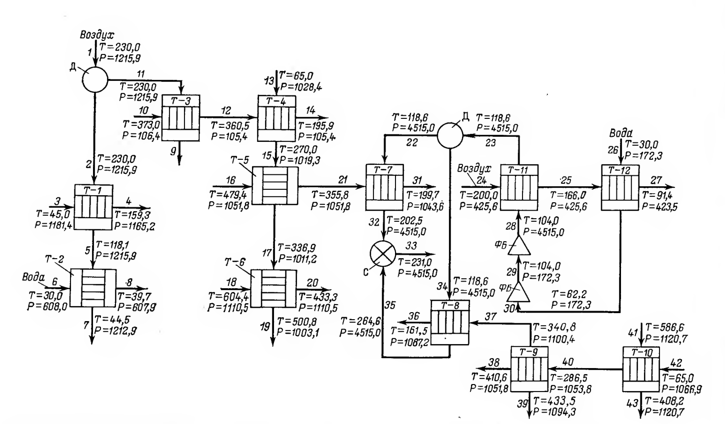


Рис. 28. Схема системы теплообмена:

1—43 — номера потоков; Т — теплообменники; Д — делители потоков; С — смеситель; ФБ — фиктивные блоки, отражающие изменение давлений и температур в аппаратах схемы, в которую теплообменная система входит как подсистема.

передачи от схемы взаимного движения теплоносителей [104];

$$\omega = \frac{G_{\text{хол}} c_{p\text{хол}}}{G_{\text{гор}} c_{p\text{гор}}} \leqslant 1$$

(G и c_p — расход и теплоемкость потока соответственно) параметр $D = kF_i/(G_{\text{хол}} c_{p\,\text{хол}})$. Поверхность теплопередачи $F_i = \pi d_i l_i N_i$ (d_i — внутренний диаметр труб; l_i — высота труб; N_i — их число; i — номер теплообменника). Коэффициент теплопередачи является функцией геометрических характеристик, в частности, числа труб N_i, их длины l_i, и шага между ними t_i. Именно эти характеристики были выбраны в качестве оптимизируемых в рассматриваемой задаче.

Таблица 29. **Начальные значения параметров технологических потоков ТС**

№ потока (см. рис. 28)	Расход м/ч *	Давление, кПа·0,01	Температура, °С	Концентрации, % (объемн.)				
				H_2O	N_2	O_2	SO_2	SO_3
1	192000	12,00	230,00	0	79,000	21,000	0	0
3	162900	11,66	45,0	1,0000	78,202	20,798	0	0
6	273009	6,00	30,0	100,00	0	0	0	0
10	162400	1,05	373,0	2,7100	92,475	3,5500	1,2650	0
13	132800	10,15	65,0	0	96,855	3,1387	0,0059	0,0004
16	133400	10,38	479,4	0	96,460	3,1250	0,0059	0,4091
18	154100	10,96	604,4	0	83,473	3,5650	1,6850	11,277
24	194000	4,20	200,0	0	79,000	21,000	0	0
26	98998	1,70	30,0	100,0	0	0	0	0
28 **	98998	46,000	104,0	100,00	0	0	0	0
39	153100	10,80	504,7	0	84,023	2,8980	0,34900	12,730
41	158800	11,06	571,2	0	81,020	6,5500	7,6600	4,77
42	133700	10,530	65,0	0	96,265	3,3210	0,39940	0,0146

* При нормальных условиях.
** Задаются начальные приближения.

Таблица 30. **Характеристики потоков ТС**

Параметр (см. рис. 28)	Тип ограничения	Заданное значение, °С	Достигнутое значение, °С	Параметр (см. рис. 28)	Тип ограничения	Заданное значение, °С	Достигнутое значение, °С
T_4	=	150	151,8	T_{27}	=	60,0	60,2
T_7	=	40	48,4	T_{43}	=	445	441,3
T_9	=	302	299,3	T_{38}	=	465	458,4
T_{14}	=	206	209,9	T_{36}	=	200	196,8
T_{20}	=	468,3	468,0	Δp (I)	$<$	1,11 *	1,1 *
T_{19}	=	529,5	533,4	Δp (II)	$<$	0,2 *	0,03 *
T_{31}	=	200,0	199,6	Δp (III)	$<$	0,03 *	0,01 *

* В кПа·10^{-2}.

С учетом всех технологических ограничений задача выбора оптимальных конструкционных параметров формулируется следующим образом. Найти значения N_i, l_i, t_i, обеспечивающие минимум функции, характеризующей общую поверхность теплообмена для всех аппаратов в схеме

$$R(N_i,\ l_i,\ t_i) = \sum_{i=1}^{12} \pi d_i l_i N_i \rightarrow \min \qquad \text{(IV, 155)}$$

при выполнении ограничений типа равенств:

$$T_i = T_i^* \qquad i = 15,\ 19,\ 21,\ 22,\ 25,\ 26,\ 29,\ 35,\ 37,\ 41,\ 43 \qquad \text{(IV, 156)}$$

и типа неравенств:

$$\Delta p\,(i) < \Delta p^*\,(i) \qquad i = \text{I, II, III} \qquad \text{(IV, 157)}$$

Перепад давления в i-том теплообменнике вычисляется по общей формуле вида:

$$\Delta p\,(i) = f\,(N_i,\ l_i,\ t_i)\,\lambda\,\frac{w^2\rho}{2} \qquad \text{(IV, 158)}$$

Здесь f — функция геометрических параметров, различно определенная для трубного или межтрубного пространства теплообменников; λ — коэффициент гидравлического сопротивления, зависящий от критерия Рейнольдса; w — линейная скорость; ρ — плотность. Переменные N_i, l_i, t_i были ограничены снизу и сверху для каждого теплообменника.

Для расчета целевой функции (IV, 155) и ограничений (IV, 156), (IV, 157) при заданной комбинации оптимизируемых переменных необходимо рассчитать стационарный режим всей схемы, показанной на рис. 28. Схема содержит один замкнутый контур, что делает необходимым проведение итераций по параметрам одного разрываемого потока. Для решеиия поставленной задачи оптимизации была использована система автоматизированного моделироваиия химико-технологических схем [105, 106], в которую включена программа оптимизации, использующая для учета ограиичеиий модифицированную функцию Лаграижа.

Решение с требуемой точиостью было получеио за 15 итераций. При этом число обращений к подпрограмме вычисления градиеита составило 16 (градиеит вычислялся по коиечным разиостям), а число обращеиий к подпрограмме расчета целевой функции (помимо вычислеиия градиента) равнялось 49. На рис. 28 указаны значения температур и давления для всех рассчитаиных потоков. Исходиые даниые приведеиы в табл. 29. Все иакладываемые ограиичения (IV, 156), (IV, 157) и достигнутые значения по ограничениям приведены в табл. 30.

В результате решения задачи оптимизации найдены значения N_i, l_i, t_i для всех теплообменников. Найденный минимум критерия составляет 5738,1.

Таким образом в результате решения задачи оптимизации по 36 переменным при наличии 14 ограничений найдены оптимальные конструкционные параметры для 12 теплообменных аппаратов.

ГЛАВА V

Оптимизация больших систем

Усложнение задач моделирования приводит к существенному увеличению их размерности. Так, в монографии [14, с. 239] утверждается, что в ближайшем будущем постановка задач исследования реальных химико-технологических схем может потребовать моделирования систем с несколькими тысячами переменных.

Как будет показано, синтез ХТС также приводит к задачам оптимизации большой размерности. Это определяет необходимость создания таких методов оптимизации, эффективность которых мало зависит от размерности задачи. В настоящее время в работах по оптимизации больших систем сформировались два направления.

Первый подход состоит в том, что схему рассматривают как единое целое и пользуются поисковыми методами оптимизации. Проанализируем в связи с этим перспективы применения поисковых методов для оптимизации больших систем. Ранее вследствие трудностей получения аналитических выражений для производных часто применялись методы поиска нулевого порядка [11, с. 121], не требующие вычисления производных. В настоящее время [107] общепринятым является использование квазиньютоновских методов первого порядка, причем в случае трудности получения аналитических выражений для производных используются их разностные аппроксимации. Однако, способ вычисления производных с помощью разностей имеет большие недостатки. Действительно, вычисление производных с помощью разностей потребует $(\bar{r} + 1)$-го расчета схемы ($\bar{r}$ — размерность вектора поисковых переменных), т. е. вычислительные затраты на определение производных в этом случае, растут пропорционально размерности задачи, и при больших $\bar{r}$ могут стать чрезмерными. Следующий недостаток — неточность расчета производных, которая может существенно исказить направления поиска, а следовательно, понизить эффективность метода. И, наконец, еще один недостаток — трудоемкость подбора приращений аргументов Δx_i.

Все указанные недостатки делают малоперспективным использование разностных оценок производных при оптимизации больших систем. Поэтому для больших систем чрезвычайно важную роль приобретают алгоритмические методы вычисления производных. От наличия таких методов и основанных на них программ будет зависеть — станут ли квазиньютоновские методы первого порядка действительно инструментом решения задач оптимизации большой размерности. Причем программы, основанные на алгоритмических методах, должны обеспечивать не только точность и быстродействие вычислений производных, но и требовать подготовительной работы

(аналитических выкладок, программирования), не намного большей, чем при использовании методов оптимизации нулевого порядка.

Метод «сопряженного процесса», позволяющий эффективно вычислять частные производные критерия [108], подробно изложен в написанной совместно с Ю. М. Волиным главе V монографии [11, с. 201]. При фиксированном числе блоков схемы вычислительные затраты этого метода мало зависят от размерности задачи оптимизации. С использованием этого метода была разработана [3, с. 267—288] система программ моделирования ХТС; для схем произвольной структуры она позволяет вычислять значения производных критерия по поисковым переменным только на основе знания математических моделей отдельных блоков, матриц Якоби правых частей соотношений (I,1) и информации о структуре ХТС.

Вместе с тем даже при «точном» вычислении производных с помощью алгоритмических методов, эффективность квазиньютоновских методов оптимизации с возрастанием $\bar{r}$, вообще говоря, должна снижаться. Так, если рассмотреть задачу минимизации выпуклой квадратичной функции для $\bar{r} = 2$ и $\bar{r} = 100$, то достижение минимума потребует в первом случае двух шагов, а во втором 100 шагов. Нет никаких оснований предполагать, что для неквадратичных функций положение изменится в лучшую сторону. Конечно, это не значит, что с увеличением $\bar{r}$ скорость сходимости всегда будет существенно уменьшаться. Решение ряда задач об оптимальном распределении параметров в реакторах идеального вытеснения и оптимизации каскада реакторов идеального смешения показали [11, с. 50, с. 91], что в упомянутых задачах при возрастании $\bar{r}$ время, затрачиваемое на решение, увеличивалось незначительно. Тем не менее, в основном следует ожидать значительного снижения эффективности квазиньютоновских методов с увеличением размерности задач. Поэтому большее значение должны приобрести метод Ньютона и его модификации, которые, по-видимому, в наибольшей степени отвечают сформулированному выше требованию о малой зависимости эффективности методов оптимизации от размерности задачи. Действительно, для выпуклой квадратичной функции независимо от ее размерности метод Ньютона должен дать решение за один шаг. Отсюда можно полагать, что и для произвольных функций, близких к квадратичным, скорость сходимости метода Ньютона будет мало зависеть от размерности задачи. Правда, при этом необходимо преодолеть затруднения [11, с. 268]. Первое и самое большое из них — нахождение матрицы вторых производных. Если уже получение первых производных с помощью разностей вызывает серьезные затруднения, то путь вычисления вторых производных разностными методами вряд ли можно рассматривать как реальный способ решения этой задачи. По-видимому, единственный путь состоит в развитии алгоритмических методов получения матрицы вторых производных целевой функции. Такой подход, основанный на использовании аппарата «сопряженного процесса», был развит в работе [109]. Второе затруднение возникает при решении систем линейных уравнений (или обращение матриц) большой размерности. Правда, как правило, матрицы этих систем являются

сильно разреженными, что делает целесообразным применение специальных методов решения таких систем [36].

Оценивая перспективы применения метода Ньютона, следует отметить, что его широкое практическое использование начнется лишь после того, как на основе развитых алгоритмических методов будут созданы программы для ЭВМ, позволяющие для схем произвольной структуры вычислять значения вторых производных критерия по поисковым переменным только на основе знания математических моделей отдельных блоков, и информации о структуре ХТС, т. е. программы, аналогичные вышеупомянутым программам вычисления первых производных. Поскольку трудно предположить, что такие программы будут созданы в ближайшие годы, основное применение найдут квазиньютоновские методы первого порядка. Как мы уже отмечали, эффективность этих методов с увеличением размерности задач должна уменьшаться. Однако, есть обстоятельство, которое позволяет существенно повысить эффективность квазиньютоновских методов: при оптимизации больших систем либо сама структура ХТС приводит к тому, что гессиан целевой функции имеет сильно разреженную структуру (большое число нулевых элементов), либо же с помощью специального приема удается получить модифицированный критерий, гессиан которого будет иметь сильно разреженную структуру. В связи с этим рассмотрим квазиньютоновские методы минимизации функций, имеющих сильно разреженные гессианы. Развитие этих методов началось в самое последнее время. Также как и в главе III мы здесь рассмотрим квазиньютоновские методы 1-го и 2-го рода.

В т о р о й п о д х о д к оптимизации больших систем состоит в использовании декомпозиционных методов оптимизации.

Декомпозиционные методы оптимизации

Декомпозиционные методы оптимизации позволяют свести глобальную задачу оптимизации ХТС большой размерности к последовательности локальных задач оптимизации отдельных блоков или совокупностей отдельных блоков («суперблоков») существенно меньшей размерности. При конструировании подобных методов главная проблема состоит в ликвидации или учете взаимного влияния блоков ХТС при формировании локальных задач оптимизации. В связи с этим был разработан принцип закрепления [110, с. 302—308], на основе которого был создан декомпозиционный метод закрепления [1, с. 302; 3, с. 182]. Независимо подобный метод был развит в работе [111]. Был предложен [112] и частный случай метода закрепления применительно к решению систем уравнений, выражающих необходимые условия оптимальности ХТС (feasible method).

Другой способ построения декомпозиционных методов основан на использовании алгоритмов сведения задач на условный экстремум к задачам на безусловный экстремум. Перечисленные подходы были подробно рассмотрены в книгах [3, с. 182; 11, с. 242]. Здесь мы коротко опищем только метод закрепления.

Воспользуемся подходом, при котором задача оптимизации ХТС сводится к задаче 3 [см. выражения (I, 75), (I, 76)], причем в качестве независимых используем переменные x', u, а в качестве зависимых — переменные z'. Для решения применим двухуровневый метод. На первом (нижнем) уровне критерий F будем оптимизировать по переменным u, на втором (верхнем) уровне — по переменным x'; ограничения (I, 76) будем учитывать на нижнем уровне. Тогда задача 3 может быть представлена в виде:

$$\min_{x',\, u \in D_3} \overline{F}(x',\, u) = \min_{x'} \min_{u \in D_3} \overline{F}(x',\, u) \tag{V, 1}$$

Обозначим через $\overline{F}'$ результат оптимизации на нижнем уровне. Ясно, что $\overline{F}'$ является функцией переменных x'

$$F'(x') = \min_{u \in D_3} \overline{F}(x',\, u) \tag{V, 2}$$

Поскольку на первом уровне все промежуточные входные переменные x', а следовательно, и промежуточные выходные переменные z' оказываются фиксированными, то варьирование переменных $u^{(k)}$, относящихся к k-му блоку, будет влиять только на часть критерия $F^{(k)}$, соответствующую k-му блоку; таким образом, на первом уровне взаимовлияния блоков ликвидируются, и задача оптимизации глобального критерия разбивается на N локальных задач оптимизации отдельных блоков с локальными критериями $F^{(k)}$ и локальными ограничениями, т. е. имеют место соотношения:

$$\overline{F}'(x') = \min_{u \in D_3} \overline{F} = \sum_{k=1}^{N} \min_{u^{(k)} \in D^{(k)}} \overline{F}^{(k)}(x',\, u)$$

где

$$D^{(k)} = \{u^{(k)}: c^{(k)} \leqslant u^{(k)} \leqslant b^{(k)};\ x^{(k)} - f^{(k)}(x^{(k)}, u^{(k)}) = 0,\ a - f^{(k)}(x^{(k)}, u^{(k)}) = 0\}$$

Этот метод можно применять, если в каждом блоке выполняется условие $r_i^{(k)} \geqslant n^{(k)}$. Характеризуя этот метод на базе сформулированных в гл. IV пяти критериев и сравнивая его с методом, при котором к схеме подходят как к единому целому, можно сказать следующее.

1. При расчете критерия отсутствует итерационная процедура.

2. При решении локальных задач всегда необходимо использовать методы условной минимизации.

3. На втором уровне необходимо решать задачу оптимизации, имеющую размерность m, (причем на каждом шаге этого уровня надо решать N локальных задач оптимизации, первого уровня, размерность каждой из которых равна r_k.

4. Учет ограничения (I, 10) не должен вносить осложнений, поскольку метод «штрафа» используется независимо от этого, а их учет приведет только к увеличению штрафных членов штрафной функции k-го блока.

5. N локальных задач нижнего уровня могут решаться независимо, что удобно для параллельного счета.

Характеристика квазиньютоновских методов минимизации функций с разреженными гессианами

Вначале покажем возможность появления критериев с сильно разреженными гессианами. Иногда сама структура ХТС приводит к задаче, в которой критерий имеет сильно разреженный гессиан. Рассмотрим в виде примера задачу оптимизации последовательно-параллельной схемы (рис. 29), имеющей l параллельных ветвей, каждая из которых содержит n блоков. Пусть номер N, $(N = m + ln + 1)$ соответствует блоку смешения. Для простоты предположим, что в каждом блоке имеется только одно управление и, следовательно, число управлений равно $m + nl$. Пусть критерий оптимизации имеет вид:

$$F = \sum_{k=1}^{N-1} F_1^{(k)}(u^{(k)}) + \sum_{k=m+n}^{m+nl} F_2^{(k)}(z^{(k)}) = \sum_{k=1}^{N} f^{(k)}(z^{(k)}) \qquad \text{(V, 3)}$$

$$f^{(k)} = \delta_{1k} \sum_{j=1}^{m} F_1^{(j)}(u^{(j)}) + \sum_{j=m+(k-1)n+1}^{m+kn} F_1^{(j)}(u^{(j)}) + F_2^{(m+kn)}(z^{(m+kn)})$$

где δ_{1k} — символ Кронекера.

Ясно, что от управлений $u^{(i)}$, $(i = \overline{1, m})$, относящихся к первым m блокам, зависят выходные переменные всех ветвей схемы. Поэтому все $f^{(k)}$ являются сложными функциями от $u^{(i)}$, $(i = \overline{1, m})$. В то же время от управлений $u^{(i)}$, $(i = m + (k-1)n + 1, \ldots, m + kn)$, относящихся к k-той ветви, зависит только совокупность выходных переменных k-той ветви, т. е. только функция $f^{(k)}$. Поэтому можно записать $f^{(k)}$ в виде:

$$f^{(k)} = f^{(k)}(u^{(1)}, \ldots, u^{(m)}, u^{(m+(k-1)n+1)}, \ldots, u^{(m+kn)})$$

Первые m строк (столбцов) гессиана функции F (рис. 30) соответствуют дифференцированию по переменным $u^{(i)}$, $(i = \overline{1, m})$, следующие n строк (столбцов) — дифференцированию по переменным $u^{(i)}$, $(i = m + 1, \ldots, m + n)$ и т. д. Здесь и далее заштрихованная часть гессиана соответствует ненулевым элементам гессиана, элементы же, стоящие вне заштрихованных частей, тождественно равны нулю. К сожалению, в большинстве случаев при оптимизации ХТС гессиан критерия не получается разреженным. Так, если в схему, показанную на рис. 29, ввести рецикл, то гессиан будет плотно заполненным

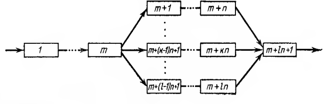


Рис. 29. Последовательно-параллельная схема.

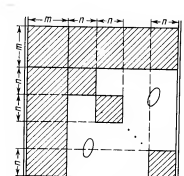


Рис. 30. Вид гессиана критерия оптимизации (V,3) для схемы, приведенной на рис. 29.

В связи с этим потребуются специальные меры, для получения разреженного гессиана. Воспользуемся подходом, при котором оптимизация ХТС сводится к задаче 1 [см. соотношения (I, 64)—(I, 66)]. В этом случае число поисковых переменных равно $\bar{r}$ [см. выражение (I,52)]. Для учета ограничений на выходные переменные применим один из методов последовательной безусловной минимизации (для определенности — метод «штрафа»). Тогда модифицированный критерий будет иметь вид:

$$\Phi_1 = \sum_{k=1}^{N} \left(F^{(k)} \ (x^{(k)}, \ u^{(k)}, \ z^{(k)}) + \beta \sum_{i=1}^{g_k} \left(z_i^{(k)} - a_i^{(k)} \right)^2 \right) \qquad (V, 4)$$

где β — штрафной коэффициент.

При такой постановке задачи гессиан критерия Φ_1 в общем случае является плотно заполненным. Чтобы получить критерий с сильно разреженным гессианом поступим следующим образом. Выберем в качестве поисковых переменных наряду с $u^{(k)}$, $(k = \overline{1, N})$ все входные переменные $x^{(k)}$, $(k = \overline{1, N})$ блоков схемы, а соотношения связи (I, 6) будем трактовать как ограничения типа равенств. В этом случае число поисковых переменных R равно

$$R = \sum_{k=1}^{N} (r_k + n_k)$$

Обычно, $R \gg \bar{r}$. Сформируем модифицированный критерий, используя один из методов сведения задач на условный экстремум к задачам на безусловный экстремум (для определенности — метод «штрафа»)

$$\Phi_2 = \sum_{k=1}^{N} \Phi_2^{(k)} \qquad (V, 5)$$

где

$$\Phi_2^{(k)} = F^{(k)} \left(x^{(k)}, u^{(k)}, z^{(k)} \right) + \beta \left[\sum_{i=1}^{g_k} \left(z_i^{(k)} - a_i^{(k)} \right) + \right.$$

$$\left. + \sum_{i=s_k+1}^{n_k} \left(x_i^{(k)} - \sum_{j=1}^{N} \alpha_{kj} f_i^{(j)} (x^j, u^{(j)}) \right) \right]^2$$

В таком виде критерий Φ_2 имеет сильно разреженный гессиан, если каждый блок ХТС связан с небольшим числом других блоков. В значительном большинстве случаев выходной поток блока не подается

на свой вход и среди его выходных переменных нет выходных переменных схемы ($q_k = 0$). При этом $\Phi_2^{(k)}$ имеет вид:

$$\Phi_2^{(k)} = F^{(k)} + \beta \sum_{i=1}^{n_k} \left[x_i^{(k)} - \sum_{j=1,\ (j \neq k)}^{N} \alpha_{kj} f^{(j)} \left(x^{(j)}, u^{(j)}\right) \right]^2 \qquad (V, 6)$$

В таком случае часть элементов гессиана функции $\Phi_2^{(k)}$ легко может быть подсчитана

$$\frac{\partial^2 \Phi_2^{(k)}}{\partial \left(x_i^{(k)}\right)^2} = \frac{\partial^2 F^{(k)}}{\partial \left(x_i^{(k)}\right)^2} + 2\beta \qquad \frac{\partial^2 \Phi_2^{(k)}}{\partial x_i^{(k)} \partial u_j^{(l)}} = -\alpha_{kl}\beta \frac{\partial f^{(l)}}{\partial u_j^{(l)}}$$

$$\frac{\partial^2 \Phi_2^{(k)}}{\partial x_i^{(k)} \partial x_j^{(l)}} = -\alpha_{kl}\beta \frac{\partial f^{(l)}}{\partial x_j^{(l)}} \qquad (V, 7)$$

Таким образом, благодаря специальному выбору поисковых переменных и существенному увеличению размерности задачи удалось привести критерий к виду, имеющему сильно разреженный гессиан. При этом существенно увеличилось число штрафных членов функции Φ_2 по сравнению с функцией Φ_1, что должно ухудшить сходимость поисковых методов. Однако, вычислительный опыт показывает, что если в критерии имеются штрафные члены, то добавление новых часто ненамного ухудшает сходимость применяемого метода. В любом случае, только опыт может дать ответ на вопрос: компенсирует ли получение сильной разреженности гессиана отрицательные последствия увеличения размерности задачи и числа штрафных членов.

В виде примера рассмотрим задачу оптимизации простой последовательности блоков (см. рис. 22). Пусть ограничения на выходные переменные блоков имеются только в N-м блоке, а критерий оптимизации имеет вид (I, 15), где $g_k = 0$, $(k = \overline{1,\ N-1})$. Соотношения связи в данном случае имеют вид (IV, 60). В отличие от соотношения (V, 5) определим функцию Φ_2 в виде:

$$\Phi_2 = \sum_{k=2}^{N+1} \Phi_2^{(k)} \qquad (V, 8)$$

где

$$\Phi_2^{(k)} = F^{(k-1)} + \beta \sum_{i=1}^{n} \left[x_i^{(k)} - f_i^{(k-1)} \left(x^{(k-1)}, u^{(k-1)}\right)\right]^2 \qquad k = \overline{2,\ N}$$

$$\Phi_2^{(N+1)} = F^{(N)} + \sum_{i=1}^{g_N} \left(z_i^{(N)} - a_i^{(N)}\right)^2$$

Таким образом, функция $\Phi_2^{(k)}$, $(k = \overline{2,\ N})$ зависит только от переменных $x^{(k-1)}$, $u^{(k-1)}$, $x^{(k)}$, а $\Phi_2^{(N+1)}$ от переменных $x^{(N)}$, $u^{(N)}$. Гессиан функции Φ_2 размерности $N\ (r + n) \times N\ (r + n)$ приведен на рис. 31. В нем первые n строк (столбцов) соответствуют дифференцированию по переменным $x^{(1)}$, следующие r строк (столбцов) — дифференцированию по переменным $u^{(1)}$, следующие n строк (столбцов) — диффе-

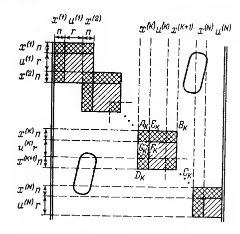


Рис. 31. Вид гессиана критерия оптимизации (I,15) для схемы, приведенной на рис. 22.

ренцированию по $x^{(2)}$, и т. д. На рис. 31 квадрат $A_kB_kC_kD_k$ соответствует гессиану $G^{(k)}$ функции $\Phi_2^{(k)}$. Поскольку для блоков с номерами $i = 1, 2, \ldots, N-1$ критерий имеет вид (V,6), часть элементов гессиана функции $\Phi_2^{(k)}$ может быть подсчитана с помощью формул (V, 7) (на рисунке эта часть отмечена двойной штриховкой). Элементы квадрата $A_kE_kF_kG_k$ соответствуют переменным $x^{(k)}$, входящим как в функцию $\Phi_2^{(k)}$, так и в функцию $\Phi_2^{(k+2)}$.

Резюмируя приведенное рассмотрение, отметим, что в гессиане могут быть элементы двух видов. Элементы первого вида — постоянные числа (в частном случае равные нулю), второго вида — легко вычисляемые [например, с помощью формул (V, 7)] выражения. Относительно построения квазиньютоновских методов минимизации функций с разреженными гессианами можно сказать то же, что было сказано о построении методов решения систем нелинейных уравнений с разреженными матрицами Якоби. Ясно, что мы должны аппроксимировать сам гессиан, а не обратную ему матрицу, поскольку гессиан может иметь большое число нулей, а его обратная матрица — быть плотно заполненной. При построении матриц B_i, аппроксимирующих гессиан G, желательно сохранить структуру самого гессиана, т. е. обеспечить равенство постоянных (в частности нулевых) и легко вычисляемых элементов матрицы G соответствующим элементам матрицы B.

Квазиньютоновские методы 1-го рода для минимизации функций с разреженным гессианом

Пусть требуется минимизировать функцию $f(x)$, где x — n-вектор с разреженным гессианом. Построение алгоритма минимизации таких функций будет основываться на подходах, развитых при выводе квазиньютоновских методов: метода 1-го рода для минимизации произвольных функций и метода решения систем нелинейных уравнений с разреженной матрицей Якоби. Так же, как и в последнем методе введем множества M_1, M_2, M, $\overline{M}$ следующим образом: M_1 — множества пар целых чисел (i, j) таких, что соответствующие элементы G_{ij} гессиана являются постоянными, не зависящими от точки (номера итерации), т. е. выполняются равенства

$$G_{ij} = p_{ij} \qquad (i, j) \in M_1 \tag{V, 9}$$

где p_{ij} — некоторые постоянные (в частном случае равные нулю); M_2 — множество пар (i, j) таких, что соответствующие элементы G_{ij} легко вычисляются;

$$M = M_1 \cup M_2 \qquad \overline{M} = \{(i, j):(i, j) \notin M\}$$

Задача состоит в построении такой матрицы B_{k+1} в $(k+1)$-й точке, которая являлась бы некоторым приближением к гессиану G_{k+1}, а ее элементы $b_{ij}^{(k+1)}$, $(k = 1, n)$ удовлетворяли бы соотношениям

$$b_{ij}^{(k+1)} = \begin{cases} p_{ij} & (i, j) \in M_1 \\ G_{ij}^{(k+1)} & (i, j) \in M_2 \end{cases} \qquad \text{(V, 10)}$$

Матрицу B_{k+1} будем искать в виде (II, 38); она должна удовлетворять квазиньютоновскому условию 1-го рода (II, 25) и условию симметричности (III, 47). Отсюда следует, что матрица E должна удовлетворять условию симметричности, а ее элементы — соотношениям

$$e_{ij} = \begin{cases} 0 & (i, j) \in M_1 \qquad \text{(V, 11)} \\ G_{ij}^{(k+1)} - G_{ij}^{(k)} = \Delta G_{ij}^{(k)} & (i, j) \in M_2 \qquad \text{(V, 12)} \end{cases}$$

Для определения матрицы B_{k+1} применим принцип наименьшего изменения матрицы B_k. Тогда задача определения матрицы E будет выглядеть следующим образом:

$$\min_{e_{ij}} \frac{1}{8} \| E \|_F \qquad \text{(V, 13)}$$

$$\sum_{j=1}^{n} e_{ij} s_j = r_{i,k} \qquad i = \overline{1, n} \qquad \text{(V, 14)}$$

$$e_{ij} = e_{ji} \qquad i, j = \overline{1, n} \qquad \text{(V, 15)}$$

$$e_{ij} = 0 \qquad (i, j) \in M_1 \qquad \text{(V, 16)}$$

$$e_{ij} = \Delta G_{ij}^{(k)} \qquad (i, j) \in M_2 \qquad \text{(V, 17)}$$

Сформулированная задача отличается от задач (III, 52)—(III, 54) и (II, 164)—(II, 167) наличием условий (V, 16), (V, 17) в первом случае и условия симметричности (V, 15) — во втором. Рассмотрим вначале случай, когда множество M_2 пусто [113]. Введем вектор-строку $s(i)$ с элементами $s(i)_j$ [см. выражение (II, 168)]. Подставим в уравнение (V, 14) значения e_{ij}, удовлетворяющие условию (V, 16). Полученные соотношения с помощью вектора $s(i)$ могут быть записаны по аналогии с уравнениями (II, 170) в виде

$$\sum_{j=1}^{n} e_{ij} s(i)_j = r_{i,k} \qquad i = \overline{1, n} \qquad \text{(V, 18)}$$

Как будет видно из дальнейшего, запись условий (V, 14) в виде (V, 18) обеспечит автоматическое выполнение условий (V, 16), поэтому соотношения (V, 18) эквивалентны соотношениям (V, 14), (V, 16). Чтобы удовлетворить условию симметричности (V, 15) будем так же, как и при решении задачи (III, 52)—(III, 54) искать матрицу E в виде (III, 56). Заменяя в задаче (V, 13)—(V, 17) соотношения (V, 14), (V, 16)

соотношением (V, 18), представляя матрицу E в виде (III, 56) и учитывая, что $M_2 = \varnothing$, сведем эту задачу к следующей

$$\min_{c_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} + c_{ji})^2$$

$$\sum_{j=1}^{n} (c_{ij} + c_{ji})\, s(i)_j = 2r_{i,\,k} \tag{V, 19}$$

Функция Лагранжа для этой задачи имеет вид

$$L = \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} + c_{ji})^2 + \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} (c_{ij} + c_{ji})\, s(i)_j$$

Приравнивая производные от L по элементам c_{ij}, нулю, по аналогии с формулой (III,57) получим

$$e_{ij} = -[\lambda_i s(i)_j + \lambda_j s(j)_i] \tag{V, 20}$$

Полученное выражение для e_{ij} удовлетворяет условию (V, 16). Действительно, из условий симметричности матрицы G следует, что если $s(i)_j = 0$, то и $s(j)_i = 0$. Отсюда, если $(i,\, j) \in M_1$, то $s(i)_j = s(j)_j$ и, следовательно, $e_{ij} = 0$. Подставив это выражение для e_{ij} в условия (V, 18), получим систему линейных уравнений для определения неизвестных величин

$$Q\lambda = r_k \tag{V, 21}$$

где элементы Q_{ij}, матрицы Q имеют вид

$$Q_{ij} = s(i)_j\, s(j)_i + \sum_{k=1}^{n} [s(i)_k]^2\, \delta_{ij}$$

а δ_{ij} — символ Кронекера. Если при решении задачи (III, 52)—(III, 54) удалось в явном виде получить выражение для E [см. выражение (III, 109)], то в данном случае для определения этой матрицы придется решать систему линейных уравнений (V, 21). Правда, эта система обладает двумя свойствами, облегчающими ее решение. Прежде всего, ее матрица Q имеет практически ту же степень разреженности, что и сам гессиан G. Действительно, если $G_{ij} = 0$ $(i \neq j)$, то $s(i)_j = 0$ и, следовательно, $Q_{ij} = 0$. Кроме того, если все векторы $s(i)$ ненулевые, то матрица Q положительно определенна [113].

Рассмотрим теперь случай, когда множество M_2 не пусто. Выше было показано, что соотношения (II,165)—(II,167) эквивалентны соотношению (II, 180). Поскольку соотношения (V, 14), (V, 16), (V, 17) совпадают с соотношениями (II,165)—(II,167) [за исключением того, что в соотношении (II,167) величины $\Delta j_{ij}^{(k)}$ надо заменить на $\Delta G_{ij}^{(k)}$], соотношения (II, 180) будут верны и в данном случае. Единственное отличие будет состоять в том, что элементы e_{ij}, $(i,\, j) \in M_2$ в величине $\bar{r}_i$ будут определяться формулами (V,12). Заменяя в задаче (V, 13)—(V, 17) соотношения (V, 14), (V, 16), (V, 17) соотношением (II, 180), представляя элементы $\bar{e}_{ij}$ в виде $\bar{e}_{ij} = 0{,}5\, (\bar{c}_{ij} + \bar{c}_{ji})$, что

гарантирует выполнение условия (V, 15), сведем эту задачу к следующей

$$\min_{c_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} (\bar{c}_{ij} + \bar{c}_{ji})^2$$

$$\sum_{j=1}^{n} (\bar{c}_{ji} + \bar{c}_{ij})\, s(i)_j = 2\bar{r}_{i,\,k} \qquad (V, 22)$$

Если в задаче (V,22) вместо $\bar{c}_{ij}$, $s(i)_j$, $\bar{r}_{i,k}$ подставить величины c_{ij}, $s(i)_j$, $r_{i,k}$, то придем к задаче (V,19). Поэтому можно воспользоваться решением задачи (V,19). Формулы для определения $\bar{e}_{ij}$ будут иметь вид (V,20), (V,21), надо только везде заменить e_{ij}, $s_i(j)_j$, $r_{i,k}$ на $\bar{e}_{ij}$, $\bar{s}(i)_j$, $\bar{r}_{i,k}$ соответственно.

При рассмотрении квазиньютоновских методов 1-го рода было показано, что лучшие результаты дает подход, при котором берется норма Фробениуса от некоторой взвешенной суммы (III, 66). Легко видеть, что в данном случае определение матрицы E сведется к решению следующей задачи

$$\min_{e_{ij}} \| WEW \| \quad (Es = r_k, \quad E^{\text{т}} = E) \qquad e_{ij} = 0 \qquad (i,\ j) \in M_1 \quad (V, 23)$$

Проведя такие же рассуждения, как при получении соотношений (III, 69), (III, 70), легко показать, что задача (V, 23) может быть сведена к следующей

$$\min_{v_{ij}} \| V \|_F \qquad V\bar{s} = r_k \qquad V^{\text{т}} = V \qquad (V, 24)$$

$$\| W^{-1}VW^{-1} \|_{ij} = 0 \qquad (i,\ j) \in M_1 \qquad (V, 25)$$

где $\bar{s}$, $\bar{r}_k$ — величины, определяемые формулами (III, 71), а через $\| W^{-1}VW^{-1} \|_{ij}$ обозначен элемент матрицы $W^{-1}EW$, стоящей на пересечении i-той строки и j-го столбца. К сожалению, в данном случае условие (V, 25) не удается так же просто учесть, как условие (V, 16) в предыдущем случае, поэтому каждое соотношение должно быть учтено в функции Лагранжа с помощью соответствующего множителя Лагранжа. В этом случае задача определения множителей Лагранжа становится трудоемкой, поскольку требует решения системы линейных уравнений большой размерности. Причем чем сильнее будет разреженность гессиана, тем больше будет условий типа (V, 25) и тем сложнее будет определение множителей Лагранжа. В связи с этим был предложен следующий подход [114]. Пусть, как и прежде, M_1 характеризует множество нулевых элементов, а M_2 — пустое множество. Вначале найдем обычным путем матрицу B, которая обеспечивает хорошую работу квазиньютоновского метода; пусть, например, это будет матрица (III, 80). Для простоты обозначим ее через B^*. Естественно, что структура матрицы G в ней не будет отражена, и, вообще говоря, она не будет содержать нулевых элементов. Поставим теперь задачу найти матрицу $\overline{B}^*$, определяемую формулой

$$\overline{B}^* = B^* + E \qquad (V, 26)$$

Элементы матрицы B^* известны, и требуется найти элементы матрицы E. Матрица $\overline{B}^*$ должна отвечать следующим условиям.

1. Удовлетворять квазиньютоновскому уравнению (II,25)

$$\overline{B}^* s_j = y_j$$

Отсюда получим условия для матрицы E, умножив формулу (V,26) на s_j, получим:

$$\overline{B}^* s_j = B^* s_j + E s_j \qquad (V, 27)$$

но поскольку по построению матрица B^* удовлетворяет условию (II, 25), равенство (V, 27) принимает вид

$$E s_j = 0 \qquad (V, 28)$$

2. Быть симметричной, поскольку таковой является матрица B^*, симметричной должна быть и матрица E.

Нулевые элементы матриц G и $\overline{B}^*$ должны соответствовать друг другу, т. е. должны выполняться соотношения

$$e_{ij} = -b^*_{ij} \qquad (i,\ j) \in M_1 \qquad (V, 29)$$

Норма Фробениуса матрицы E должна быть минимальной. Таким образом, матрица E должна быть определена решением следующей задачи:

$$\min_{e_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}^2 \qquad (V, 30)$$

$$\sum_{j=1}^{n} e_{ij} s_j = 0 \qquad i = \overline{1,\ n} \qquad e_{ij} = e_{ji} \qquad i = \overline{1,\ n}$$

$$e_{ij} = -b^*_{ij} \qquad (i,\ j) \in M_1$$

Формально задача (V,30) может быть сведена к ранее рассмотренной задаче (V,13)—(V,17), если положить $r_{i,k=0}$, множество M_1 считать пустым, множество M_2 обозначить через M_1, а величины $\Delta G_{ij}^{(k)}$ заменить на $-b^*_{ij}$. Следовательно, для решения задачи (V, 30) мы можем воспользоваться методом решения, изложенным выше.

Блочные квазиньютоновские методы 2-го рода *

Основная идея этого подхода состоит в следующем. При построении обычных квазиньютоновских методов (см. гл. III) ищут аппроксимацию гессиана для всей целевой функции. В данном же случае, когда известна структура целевой функции, вначале аппроксимируют гессианы отдельных частей целевой функции, после чего конструируют аппроксимацию к гессиану самой целевой функции [115].

Рассмотрение, приведенное выше [см. выражения (V, 3), (V, 5)] показывает, что критерий f в задачах оптимизации XTC может быть представлен в виде:

$$f(x) = \sum_{k=1}^{N} f^{(k)}(x^{(k)}) \qquad (V, 31)$$

где x — n-вектор поисковых переменных; $x^{(k)}$ — n_k-вектор поисковых переменных, от которых фактически зависит функция $f^{(k)}$. Обозначим

* Раздел написан совместно с Е. М. Михайловой.

через x_i, $x_i^{(k)}$ значения векторов x, $x^{(k)}$ на i-той итерации, через $x_{l,i}$, $x_{li}^{(k)}$ — l-ю компоненту вектора x_i, $x_i^{(k)}$, соответственно. В продолжение итерационной процедуры минимизации f необходимо строить некоторую аппроксимацию гессиана функции f. При этом справедлива формула

$$G = \sum_{k=1}^{N} G^{(k)} \tag{V, 32}$$

где G, $G^{(k)}$ — гессианы функций f, $f^{(k)}$, соответственно. Построение аппроксимации B гессиана G будет протекать в следующем порядке. Вначале на i-той итерации будут определяться аппроксимации $B_i^{(k)}$, $(k = \overline{1, N})$ гессианов $G_i^{(k)}$ $(k = \overline{1, N})$, потом по аналогии с формулой (V, 32) будет строиться матрица B_i

$$B_i = \sum_{k=1}^{N} B_i^{(k)} \tag{V, 33}$$

Рассмотрим два способа получения матрицы $B_i^{(k)}$. При первом способе используется специальный вид функций $f^{(k)}$; из выражения (V, 5) следует, что функции $f^{(k)}$ могут быть представлены в виде

$$f^{(k)} = \frac{1}{2} \sum_{l=1}^{n} \varphi_{lk} \, (x^{(k)})^2 + \Psi_k \, (x^{(k)}) \tag{V, 34}$$

где функции φ_{lk} соответствуют штрафным членам, а Ψ_k является другим обозначением функции $F^{(k)}$. Вид функций φ_{lk} соответствует выражению (V, 5). Будем исходить из предположения, что гессиан функции Ψ $(x^{(k)})$ легко может быть выписан. Эта ситуация создается в случае, когда часть критерия F, относящаяся к k-му блоку, зависит только от управляющих переменных $u^{(k)}$. Однако зависимость $F^{(k)}$ от $u^{(k)}$ обычно является достаточно простой, что позволяет выписать гессиан этой функции. В данном случае требуется аппроксимировать только сумму, стоящую в правой части равенства (V,34). Используя правило дифференцирования сложных функций, найдем

$$\frac{\partial^2 f^{(k)}}{\partial x_i \, \partial x_j} = \sum_{l=1}^{n} \frac{\partial \varphi_{lk}}{\partial x_i} \cdot \frac{\partial \varphi_{lk}}{\partial x_j} + \sum_{l=1}^{n} \varphi_{lk} \frac{\partial^2 \varphi_{lk}}{\partial x_i \, \partial x_j} + \frac{\partial^2 \Psi}{\partial x_i \, \partial x_j} \tag{V, 35}$$

Аппроксимируем гессиан функции $f^{(k)}$, отбросив в правой части этого равенства вторую сумму, т. е. определим аппроксимацию $B^{(k)}$ гессиана G^k следующим образом:

$$B^{(k)} = \sum_{l=1}^{n} \frac{\partial \varphi_{lk}}{\partial x_i} \cdot \frac{\partial \varphi_{lk}}{\partial x_j} + \frac{\partial^2 \Psi}{\partial x_i \, \partial x_j} \tag{V, 36}$$

Эта аппроксимация будет тем более точной, чем ближе функции φ_{lk} к линейным, а итерационная точка к точке минимума, поскольку в этом случае функция φ_{lk} близка к нулю. При соблюдении этих условий

вторая сумма в выражении (V, 35) будет величиной второго порядка малости по сравнению с первой суммой.

Рассмотрим теперь другой способ построения матриц $B^{(k)}$. Выпишем некоторые очевидные формулы

$$g_i = \sum_{k=1}^{N} g_i^{(k)} \tag{V, 37}$$

где g_i, $g_i^{(k)}$ — соответственно градиенты функций f, $f^{(k)}$ в i-ой точке

$$y_i = \sum_{k=1}^{N} y_i^{(k)} \tag{V, 38}$$

где

$$y_i = g_{i+1} - g_i \qquad y_i^{(k)} = g_{i+1}^{(k)} - g_i^{(k)}$$

Вначале рассмотрим случай, когда все $f^{(k)}$, а следовательно и f — квадратичные функции. При этом гессианы G, $G^{(k)}$ будут матрицами коэффициентов соответствующих квадратичных функций. Если аппроксимацию B гессиана G строить сразу для функции f, то в i-той точке B_i можно найти из матричного уравнения (II, 29). Вид матриц S_i, Y_i задается формулами (II, 27).

Перейдем теперь к нахождению аппроксимации каждого гессиана всех составляющих $f^{(k)}$ функции f. Предположим, что доступны вычислению как сама функция $f^{(k)}$, так и ее градиент $g^{(k)}$. Обозначим через $B_i^{(k)}$ некоторую аппроксимацию гессиана $G_i^{(k)}$ функции $f^{(k)}$ в точке x_i. По аналогии с соотношением (II, 29) найдем $B_i^{(k)}$ из матричного уравнения

$$B_i^{(k)} S_i = Y_i^{(k)} \tag{V, 39}$$

(где $Y_i^{(k)} = \left(y_0^{(k)}, \ \ldots, \ y_{i-1}^{(k)}\right)$ с помощью любой из формул семейства Адачи (II, 90), (II, 91) (подробнее об этом см. ниже). Пусть определены все $B_i^{(k)}$, тогда матрицу B_i будем формировать с помощью формулы (V, 33). Покажем, что матрица B_i, построенная с помощью формулы (V, 33), в которой $B_i^{(k)}$ удовлетворяют соотношениям (V,39), будет удовлетворять соотношению (II,29). Действительно, из формулы (V, 38) вытекает равенство:

$$Y_i = \sum_{k=1}^{N} Y_i^{(k)} \tag{V, 40}$$

Подставим полученное выражение для Y_i и B_i из формулы (V, 33) в соотношение (II, 29)

$$\sum_{k=1}^{N} \left(B_i^{(k)} S_i - Y_i^{(k)}\right) = 0$$

Поскольку матрицы $B_i^{(k)}$ удовлетворяют уравнениям (V, 39), мы пришли к тождеству и наше утверждение доказано. Если размерности векторов $x^{(k)}$ и x совпадают, такой подход мало что может дать. Иначе обстоит дело, когда размерности векторов $x^{(k)}$ существенно меньше размерности вектора x. Пусть выполняется условие $n_k \ll n$. Обозна-

чим через M_k множество номеров компонентов вектора $x^{(k)}$, от которых фактически зависит $f^{(k)}$

$$M_k = \{i_1, \ldots, i_{n_k}\} \qquad (V, 41)$$

а через M ($M = \{i : 1, 2, \ldots, n\}$) — множество индексов переменных, от которых зависит функция f. Рассмотрим пример. Пусть

$$f = f^{(1)}(x_1, x_2) + f^{(2)}(x_1, x_3) + f^{(3)}(x_1, x_3, x_4) \qquad (V, 42)$$

и $x^{(1)} = (x_1, x_2)$; $x^{(2)} = (x_1, x_3)$; $x^{(3)} = (x_1, x_3, x_4)$; $M_1 = \{1, 2\}$; $M_2 = \{1, 3\}$; $M_3 = \{1, 3, 4\}$. В соответствии с определением множества M_k справедливы соотношения

$$g_{li}^{(k)} = y_{li}^{(k)} = 0 \qquad l \notin M_k \qquad (V, 43)$$

$$G_{lj}^{(k)} = 0 \qquad l \notin M_k \text{ или } j \notin M_k \qquad (V, 44)$$

Поскольку $f^{(k)}$ фактически зависит от небольшого числа переменных, векторы $g^{(k)}$, $y^{(k)}$ будут иметь большое число нулевых компонент, а гессианы $G^{(k)}$ — достаточно много нулевых строк и столбцов, хранить которые не имеет смысла. Обозначим через $\bar{g}^{(k)}$ градиент функции $f^{(k)}$, компонентами которого являются только производные $f^{(k)}$ по переменным, фактически входящим в вектор $x^{(k)}$. Введем обозначения

$$\bar{y}_i^{(k)} = \bar{g}_{i+1}^{(k)} - \bar{g}_i^{(k)} \qquad \bar{s}_i^{(k)} = x_{i+1}^{(k)} - x_i^{(k)} \qquad (V, 45)$$

$$\bar{Y}_i^{(k)} = \left(\bar{y}_0^{(k)}, \ldots, \bar{y}_{i-1}^{(k)}\right) \qquad \bar{S}_i^{(k)} = \left(\bar{s}_0^{(k)}, \ldots, \bar{s}_{i-1}^{(k)}\right) \qquad i \leqslant n_k \qquad (V, 46)$$

Формально векторы $\bar{g}_i^{(k)}$, $\bar{y}_i^{(k)}$, $\bar{s}_i^{(k)}$ получают из векторов $g_i^{(k)}$, $y_i^{(k)}$, s_i, вычеркнув нулевые компоненты. Происходит «сжатие» этих векторов. Соответствие между компонентами векторов $\bar{g}_i^{(k)}$, $\bar{y}_i^{(k)}$, $\bar{s}_i^{(k)}$ и $g_i^{(k)}$, $y_i^{(k)}$, s_i можно представить с помощью следующей «таблицы соответствия»:

$$A_k = \begin{pmatrix} 1 & \ldots & l & \ldots & n_k \\ j_1 & \ldots & j_l & \ldots & j_{n_k} \end{pmatrix} \qquad (V, 47)$$

Верхняя строчка «таблицы» A_k соответствует номерам компонент векторов $\bar{g}_l^{(k)}$, $\bar{y}_l^{(k)}$, $\bar{s}_l^{(k)}$, а нижняя — номерам компонент векторов $g_l^{(k)}$, $y_l^{(k)}$, s_l. «Таблица» A_k задает соответствие между элементами верхней и нижней строки: так, элементу l-той верхней строки соответствует стоящий под ним элемент нижней строки j_l. Это соответствие означает, что l-тая компонента вектора $\bar{g}_i^{(k)}$, $\bar{y}_i^{(k)}$, $\bar{s}_i^{(k)}$ является j_l-той компонентой вектора $g_i^{(k)}$, $y_i^{(k)}$, s_i соответственно:

$$\bar{g}_{lq}^{(k)} = g_{j_l q}^{(k)} \qquad \bar{s}_{lq}^{(k)} = s_{j_l q} \qquad \bar{y}_{lq}^{(k)} = y_{j_l q}^{(k)} \qquad l = \overline{1, n_k} \qquad (V, 48)$$

Ясно, что множество M_k может быть записано в виде $M_k = \{j_l : l = 1, n_k\}$. Тогда «таблицы» A_k для соотношения (V, 42) будут иметь вид

$$A_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix} \qquad (V, 49)$$

Соответствие элементов нижней строки A_k элементам ее верхней строки и обратное соответствие будем иногда записывать в виде

$$j_l = A_k(l) \qquad l = A_k^{-1}(j_l) \tag{V, 50}$$

В качестве примера приведем ряд соотношений (V. 50) для «таблиц» (V, 49).

$$3 = A_2(2) \qquad 2 = A_2^{-1}(3) \qquad 4 = A_3(3) \qquad 3 = A_3^{-1}(4)$$

Образуем теперь гессиан функции $f^{(k)}$ при условии, что вторые производные в нем берут только по фактически входящим в $f^{(k)}$ переменным. Этот гессиан обозначим через $\overline{G}^{(k)}$

$$\overline{G}^{(k)} = \left(\frac{\partial^2 f^{(k)}}{\partial x_i\, \partial x_j}\right)_{(i,\, j) \in M_k} \tag{V, 51}$$

Формально матрицу $\overline{G}^{(k)}$ получают из $G^{(k)}$, вычеркивая все нулевые строки и столбцы. Числа нулевых строк и столбцов равны. Для соотношения (V, 42) матрицы $G^{(k)}$, $\overline{G}^{(k)}$ будут иметь вид

$$G^{(1)} = \begin{pmatrix} \times & \times & 0 & 0 \\ \times & \times & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad G^{(2)} = \begin{pmatrix} \times & 0 & \times & 0 \\ 0 & 0 & 0 & 0 \\ \times & 0 & \times & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$G^{(3)} = \begin{pmatrix} \times & 0 & \times & \times \\ 0 & 0 & 0 & 0 \\ \times & 0 & \times & \times \\ \times & 0 & \times & \times \end{pmatrix}$$

$$\overline{G}^{(1)} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix} \qquad \overline{G}^{(2)} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix} \qquad \overline{G}^{(3)} = \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}$$

Здесь значок $\times$ обозначает ненулевой элемент. Построим аппроксимацию $B_l^{(k)}$ гессиана $G_l^{(k)}$. Вначале рассмотрим случай, когда $n_1 = n_2 = \cdots = \bar{n}$. Матрицу $B_l^{(k)}$ будем строить таким образом, чтобы она сохранила структуру матрицы $G_l^{(k)}$, т. е. все нулевые строки и столбцы матрицы $G_l^{(k)}$ стали нулевыми строками и столбцами матрицы $B_l^{(k)}$. Итак, в соответствии с (V, 44) все строки матрицы $B_l^{(k)}$ с номерами $l \notin M_k$ и столбцы с номерами $i \notin M_k$ будут нулевыми, т. е.

$$b_{li}^{(k)} = 0 \qquad l \notin M_k \qquad i \notin M_k \tag{V, 52}$$

По аналогии с матрицей $\overline{G}_l^{(k)}$ введем матрицу $\overline{B}_l^{(k)}$, которую получают из матрицы $B_l^{(k)}$, вычеркивая все нулевые строчки и столбцы, т. е. строки с номерами $i \notin M_k$ и столбцы с номерами $j \notin M_k$. Элементы, стоящие на пересечении l-той строки и q-го столбца матрицы B_i, $B_l^{(k)}$, $\overline{B}_l^{(k)}$, обозначим через b_{lq}, $b_{lq}^{(k)}$, $\bar{b}_{lq}^{(k)}$, соответственно. Для про-

стоты записи опущен индекс i — номер шага. Естественно, что строчки (столбцы) матрицы $B_i^{(k)}$ в матрице $\overline{B}_i^{(k)}$ получат другую нумерацию. Соответствие между номерами строк (столбцов) матриц $\overline{B}_i^{(k)}$ и $B_i^{(k)}$ может быть задано с помощью «таблицы» A_k. В верхней ее строке располагаются номера строк (столбцов) матрицы $\overline{B}_i^{(k)}$, а в нижней — номера соответствующих строк (столбцов) матрицы $B_i^{(k)}$. Так, l-тая строка (столбец) матрицы $\overline{B}_i^{(k)}$ становится j_l-той строкой (столбцов) матрицы $B_i^{(k)}$. При этом, очевидно, выполняются соотношения

$$\overline{b}_{ls}^{(k)} = b_{j_l j_s}^{(k)} \qquad l = \overline{1,\ n_k} \qquad s = \overline{1,\ n_k} \qquad j_l \in M_k \qquad j_s \in M_k \quad (V, 53)$$

Матрицу $B_i^{(k)}$ построим следующим образом. Вначале матрицу определим $\overline{B}_i^{(k)}$. Зная матрицу $\overline{B}_i^{(k)}$, с помощью соотношений (V, 52), (V, 53) построим матрицу $B_i^{(k)}$. Матрицу $\overline{B}_i^{(k)}$ определим из матричного уравнения

$$\overline{B}_i^{(k)} \overline{S}_i^{(k)} = \overline{Y}_i^{(k)} \quad (V, 54)$$

Здесь матрицы $\overline{S}_i^{(k)}$, $\overline{Y}_i^{(k)}$ определяются соотношениями (V, 46). Покажем теперь, что если выполняются равенства (V, 54), (V, 52), то соблюдаются равенства (V, 39). Используя правило перемножения матриц, получим, что равенство (V, 39) эквивалентно in^2 соотношениям

$$\sum_{r=1}^{n} b_{lr}^{(k)} s_{rq} = y_{lq}^{(k)} \qquad l = \overline{1,\ n} \quad q = \overline{0,\ i-1} \quad (V, 55)$$

а равенство (V, 54) эквивалентно in_k^2 соотношениям

$$\sum_{r=1}^{n_k} \overline{b}_{lr}^{(k)} \overline{s}_{rq}^{(k)} = \overline{y}_{tq}^{(k)} \qquad t = \overline{1,\ n_k} \quad q = \overline{0,\ i-1} \quad (V, 56)$$

Итак, пусть выполняются соотношения (V, 56), (V, 52), тогда будут выполняться соотношения (V, 55). Действительно, при $l \notin M_k$ в соответствии с соотношениями (V, 43), (V, 52) левые и правые части выражения (V, 55) обращаются в нуль и, следовательно, это соотношение выполняется. Пусть теперь $l = j_t \in M_k$. В соответствии с условием (V, 52) соотношение (V, 55) можно переписать в виде

$$\sum_{r=1}^{n_k} b_{j_t j_r}^{(k)} s_{j_r q} = y_{j_t q}^{(k)} \quad (V, 57)$$

Подставляя в это выражение величину $b_{j_t j_r}$ из (V, 53) и величины $s_{j_r q}$, $y_{j_t q}^{(k)}$ из условий (V, 48), получим соотношение (V, 56), что и доказывает правильность соотношения (V, 55), а, следовательно, и (V, 39). Ранее было доказано, что из соотношения (V, 39) следует равенство (II, 29). Итак, получаем следующий способ построения матрицы B_i. Вначале следует найти матрицы $\overline{B}_i^{(k)}$ из систем уравнений (V, 54). После этого надо было бы сформировать матрицы $B_i^{(k)}$ ($k = \overline{1,\ N}$) и с помощью соотношения (V, 33) найти матрицу B_i. Факти-

чески, нет необходимости формировать матрицы $B_i^{(k)}$. Зная соотношения (V, 52), (V, 53), можно сразу формировать элементы матрицы B_i из элементов матриц $\overline{B}_i^{(k)}$. Действительно, используя соотношения (V, 33), (V, 52), (V, 53) и (V, 50), получим

$$b_{ij} = \sum_{k=1}^{N} b_{ij}^{(k)} \qquad (V, 58)$$

где

$$b_{ij}^{(k)} = \begin{cases} 0 & i \notin M_k \qquad j \notin M_k \\ \bar{b}_{pq}^{(k)} & p = A_k^{-1}(i) \qquad q = A_k^{-1}(j) \qquad i, j \in M_k \end{cases}$$

Пусть теперь для $i = 1, \ldots, n$ оказалось возможным определить все матрицы $\overline{B}_i^{(k)}$, $(k = \overline{1, N})$ из системы (V, 54). Тогда при $i = \bar{n}$ в соответствии с выражением (III, 109) получим $B_{\bar{n}}^{(k)} = G^{(k)}$, т. е. точное значение гессиана квадратичной функции $f^{(k)}$, а в соответствии с равенством (V, 33) и точное значение гессиана G всей квадратичной функции f. Отсюда при $i = \bar{n}+1$ найдем точку минимума (при условии, что $\det G_n \neq 0$). Для определения матрицы $\overline{B}_i^{(k)}$ из уравнения (V, 54) может быть использована любая формула из семейства (II, 90), (II, 91), при выводе которых не использовалось свойство сопряженности направлений. Нельзя использовать такие широкоизвестные формулы, как DFP, BFGS и др. Это связано со следующим обстоятельством. Ранее было доказано, что если матрица H_i удовлетворяет уравнению (II, 32), то направления p_j, $(j = 0, 1, \ldots, n-1)$, даваемые формулой (I, 41), будут сопряженными. Аналогичное утверждение справедливо, когда строят матрицу B_i. Следовательно, если только решение систем уравнений (V, 54) может быть проведено для всех $i = 1, \ldots, \bar{n}$, то направления p_i, $(i = 0, 1, \ldots, \bar{n}-1)$ в полном пространстве переменных x будут G-сопряженными. Однако, утверждать, что векторы $s_i^{(k)}$ будут $G^{(k)}$-сопряженными, нельзя, поскольку в подпространстве переменных $x^{(k)}$ направление поиска $\bar{p}_i^{(k)}$ будет определяться не формулой $\overline{B}_i^{(k)} \bar{p}_i^{(k)} = -\bar{g}_i^{(k)}$, а проекцией вектора p_i на подпространство переменных $x^{(k)}$. Отсутствие $\overline{G}^{(k)}$-сопряженности векторов $\bar{s}_i^{(k)}$, $(i = 0, 1, \ldots, \bar{n}-1)$ ведет к невозможности гарантировать их линейную независимость. При их линейной зависимости или близости к таковой возникнут трудности с определением матрицы $\overline{B}_i^{(k)}$ из системы (V, 54).

Рассмотрим теперь случай, когда величины n_k могут быть различными. Рассуждая так же, как в случае применения блочного квазиньютоновского метода расчета ХТС при различной размерности потоков, легко показать, что минимум квадратичной функции будет найден при $i = \max n_k + 1$.

Сравним теперь рассмотренный метод с обычным квазиньютоновским методом. При применении последнего для хранения матрицы H_i или B_i потребуется n^2 ячеек памяти, и минимум будет достигнут за n шагов. При использовании изложенного подхода и применении, например, формул (II, 103), (II, 104) для нахождения матриц $\overline{B}_i^{(k)}$ по-

требуется $2\left(n_1^2+\cdots+n_N^2\right)$ ячеек памяти и минимум будет найден за число шагов, равное $\max n_k + 1$. Пусть, например, $n = 100$, $N = 10$, $n_k = 10$, $(k = \overline{1,\ 10})$. Тогда обычный подход потребует 10^4 ячеек памяти и минимум будет найден за 100 шагов, при применении изложенного подхода потребуется $2 \cdot 10^3$ ячеек памяти и минимум будет найден за 11 шагов.

В качестве примера рассмотрим последовательно-параллельную схему (см. рис. 29). В этом случае функция f имеет вид (V, 3). Обычный квазиньютоновский метод потребует $(m + nl)^2$ ячеек памяти ЭВМ для хранения элементов матрицы B_i. Метод же, изложенный выше, потребует $2l\,(n + m)^2$ ячеек для хранения матрицы B_i. Если критерий (V, 3) будет квадратичной функцией переменных $u^{(k)}$, $z^{(k)}$, а модели блоков — линейными, то обычный квазиньютоновский метод потребует $m + nl$ итераций, а рассмотренный — только $m + n + 1$ итераций. Заметим, что эффект уменьшения числа итераций связан со слабой заполненностью гессианов функции $f^{(k)}$, а не гессиана самой функции f. Гессиан функции f может быть сильно заполненным, тем не менее эффект уменьшения числа итераций будет наблюдаться, если гессианы функций $f^{(k)}$ будут сильно разреженными. В этом может быть преимущество таких методов по сравнению с квазиньютоновскими методами 1-го рода, для которых существенна сильная разреженность самого гессиана функции f. Преимущество перед квазиньютоновскими методами 1-го рода состоит также в том, что блочные квазиньютоновские методы обладают свойством квадратичного окончания, т. е. они позволяют найти минимум квадратичной функции за число шагов, равное максимальной размерности векторов $x^{(k)}$. Однако, при применении данного подхода могут возникнуть и трудности, связанные с определением матрицы $\overline{B}_i^{(k)}$ из уравнения (V, 54) в случае близости к линейной зависимости n_k векторов $\bar{s}_0^{(k)}, \ldots, \bar{s}_{n_k-1}^{(k)}$. Если такая ситуация возникает, надо предпринимать специальные меры.

Проиллюстрируем рассмотренный подход для схемы, представленной на рис. 22. Функция (V, 31) будет иметь вид (V, 8). Размерность гессианов $\overline{G}^{(k)}$ равна $2n + r$. «Таблицы соответствия» A_k будут иметь вид:

$$A_k = \begin{pmatrix} 1 & \ldots & n & \ldots & n+r & \ldots & 2n+r \\ l_k+1 & \ldots & l_k+n & \ldots & l_k+n+r & \ldots & l_k+2n+r \end{pmatrix} \quad l_k = (n+r)\,(k-1)$$

Поскольку часть гессиана $\overline{G}^{(k)}$ легко вычисляется, аппроксимировать придется только часть гессиана $\overline{G}^{(k)}$, имеющую размерность $n + r$ (размерность матрицы $\overline{B}_k$ будет не $2n + r$, а $n + r$). Оценим, что может дать этот подход. Пусть модели всех блоков линейны, и $\Phi_2^{(k)}$ квадратичные функции. Тогда при использовании обычного квазиньютоновского метода потребуется $(Nr + n)^2$ ячеек памяти для хранения матрицы H_i или B_i, и минимум будет найден за $(Nr + n)$ шагов. При применении рассмотренного метода потребуется $2N\,(2n + r)^2$ ячеек памяти и минимум будет найден за $(n + r)$ шагов. Пусть, например, $N = 30$, $n = 3$, $r = 2$; тогда при использовании

обычного квазиньютоновского метода придется хранить около 4000 элементов матрицы B_i или H_i и минимум будет найден за 63 шага, а при применении рассмотренного метода минимум будет найден за 6 шагов, и необходимо будет хранить около 2000 элементов.

Примечание. Подход, в соответствии с которым аппроксимируются отдельные части гессиана, а затем конструируется полный гессиан, может быть применен и к более сложным функциям, чем выражения (V, 31). Пусть, например, минимизируемая функция имеет вид:

$$f = F(f_1, \ldots, f_p)$$

где $f_i = f_i(x_1, \ldots, x_n)$ — функции n переменных, а F — функция p переменных. Используя правило дифференцирования сложных функций, легко получить выражения для вторых производных функции f:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \sum_{k=1}^{p} \left(\frac{\partial F}{\partial f_k} \cdot \frac{\partial^2 f_k}{\partial x_i \partial x_j} + \sum_{l=1}^{p} \frac{\partial^2 F}{\partial f_k \partial f_l} \cdot \frac{\partial f_k}{\partial x_i} \cdot \frac{\partial f_l}{\partial x_j} \right) \qquad \text{(V, 59)}$$

Формула (V,59) связывает элементы гессианов функций f и f_i ($i = \overline{1, p}$). Пусть способом, описанным выше, построены аппроксимации $B^{(k)}$ гессианов функции $f^{(k)}$; тогда по аналогии с выражением (V, 59) элементы матрицы B будем подсчитывать с помощью формулы

$$b_{ij} = \sum_{k=1}^{p} \left(\frac{\partial F}{\partial f_k} b_{ij}^{(k)} + \sum_{l=1}^{p} \frac{\partial^2 F}{\partial f_k \partial f_l} \cdot \frac{\partial f_k}{\partial x_i} \cdot \frac{\partial f_l}{\partial x_j} \right)$$

Рассмотрим частный случай, когда f_i —квадратичные функции. Тогда, если мы используем первый способ для получения матриц $B^{(k)}$, то, начиная с n-го шага и далее, будем иметь точные значения гессианов функции f_i, а, следовательно, в соответствии с формулой (V, 59), — точное значение гессиана функции f; т. е. через n шагов будет работать метод Ньютона.

В качестве примера была проведена оптимизация следующих двух функций (f и φ), имеющих структуру (V, 3):

$$f = \sum_{k=1}^{10} \left(\sum_{j=1}^{3} (k + j - 1)^2 \left(u_j^2 + u_{3k+j}^2 \right) + k \right)$$

$$\varphi = f + \sum_{k=1}^{10} \left[k^2 u_1^2 u_{3k+1}^2 + (k+1)^2 u_2^3 u_{3k+2}^2 + k^2 u_3^2 u_{3k+3}^2 + \right.$$

$$\left. + (k+1)^2 u_1^2 u_{3k+1}^2 + k^2 u_2^2 u_{3k+3}^2 + (k+1)^2 u_3^2 u_{3k+1}^2 \right]$$

Для оптимизации был использован метод DFP, а также блочный квазиньютоновский метод 2-го рода (БКМ-2). Для определения матриц $\overline{B}_i^{(k)}$ из уравнения (V, 54) использовались формулы (II,103), (II,104). Поиск проводился из точки $u_i = 1, 2$; $i = 1, \ldots, 15$; $u_j = 10$; $j = 16, \ldots, 33$. Во всех случаях было получено значение целевой функции 55. Результаты счета приведены в табл. 31. Все функции имеют по 33 переменных; каждая отдельная функция $f^{(k)}$

Таблица 31. **Сравнение эффективности метода и блочного метода БКМ-2**

Обозначение		Значение функции в начальной точке	K_f	K_p	Обозначение		Значение функции в начальной точке	K_f	K_p
метода	функции				метода	функции			
DFP	f	139010	78	16	DFP	φ	487724	190	59
БКМ-2	f	139010	30	7	БКМ-2	φ	487724	101	23

имеет размерность 6. Поскольку функция f является квадратичной, теоретически решение задачи должно было бы быть получено для $k_p = 7$; фактически же решение получено для $k_p = 7$. Для неквадратичной функции φ использование метода БКМ-2 позволило уменьшить число направлений поиска и число вычислений функции в 2 раза. Интересно отметить, что минимизация функции f методом наискорейшего спуска потребовала построения 430 направлений и 1500 вычислений функции.

ГЛАВА VI

Оптимальный синтез химико-технологических схем

При создании новых химико-технологических схем всегда стремились к построению эффективных и экономичных систем. Однако эту задачу решали на основе интуиции и опыта, другими словами, на основе ряда простых эвристических правил, найденных эмпирически. В качестве примера приведем некоторые эвристические правила построения реакторных схем.

1. Если реакторный узел представляет собой последовательность реакторов идеального смешения, то, если степень превращения в нем недостаточна, необходимо увеличить число реакторов. Если же имеется реактор идеального вытеснения, то для повышения степени превращения необходимо увеличивать его длину. Однако при этом растут капитальные затраты.

2. Увеличения степени превращения можно добиться, предусмотрев рецикл (см. рис. 5). Это особенно важно, когда исходное сырье дорого и сброс непрореагировавших продуктов невыгоден. Создание малоотходных производств также требует введения рецикла. Однако это приводит к увеличению энергетических расходов на перекачку обратного потока, так что приходится искать компро-

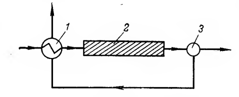


Рис. 32. Схема реактора с тепловой обратной связью:
1 — теплообменник; *2* — реактор; *3* — делитель потока.

мисс между увеличением степени превращения и ростом капитальных и энергетических затрат.

3. Тепло выходного потока реактора, в котором протекает экзотермическая реакция, может быть использовано либо для подогрева исходной смеси (рис. 32), либо для получения пара.

4. Применение исходной смеси для охлаждения промежуточных потоков в реакторе с адиабатическими слоями, в которых протекают экзотермические реакции; охлаждение может проводиться как с помощью теплообмена, так и смешения исходного и промежуточных потоков (рис. 33).

Известен ряд эвристических правил для построения схем разделения [116] и теплообменных систем [117]. Итак, даже при построении реакторной схемы мы сталкиваемся с необходимостью выбора наилучшей схемы из большого числа различных вариантов. Так, реакцию можно проводить в реакторах смешения или вытеснения либо в их комбинации, может варьироваться их число, употребляться или не употребляться рецикл, возможны различные схемы теплообмена исходного потока с промежуточными и выходными потоками реакторного узла. Выбор одного из огромного числа вариантов основывался на интуиции проектировщика. Теперь же ставится задача поручить эту творческую работу (или хотя бы ее часть) электронной вычислительной машине. Другими словами, ставится задача создания теории построения (синтеза) ХТС [116], [118], [119]. При этом возможны два пути. Первый путь — формализация того способа мышления, которым пользуется человек при создании новых схем, формализация существующих эвристических правил, создание новых, а также разработка методов использования этих правил, приоритета одних перед другими, и т. д. Второй путь — полностью алгоритмический подход, состоящий в том, чтобы сформулировать проблему синтеза как математическую и развить математические методы ее решения. Не давая окончательного ответа на вопрос, какой путь лучше, приведем пример совсем из другой области. Многовековая эволюция живого мира привела к способу передвижения живых существ с помощью ног. Многочисленные изобретения средств

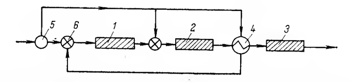


Рис. 33. Схема адиабатического реактора с охлаждением промежуточных потоков:
1—3 — слой катализатора; *4* — теплообмениик; *5* — делитель потоков; *6* — смеситель потоков.

передвижения с помощью некоторых аналогов ног, т. е. копирование живого мира, не приводило к хорошим результатам. Изобретение же человеком колеса — средства, которое не было создано эволюцией живого мира, оказалось революционным для решения проблемы создания средств передвижения. Этот пример не должен создать впечатления у читателя, что авторы являются сторонниками только второго пути, однако здесь мы будем рассматривать только подходы и методы только этого пути.

Задача синтеза ХТС может быть сформулирована следующим образом. Пусть известны сходные вещества X_i, $(i = \overline{1, p})$ и вещества Y_i, $(i = \overline{1, m})$, которые требуется получить, а также имеется набор из N аппаратов, или блоков (реакторы, ректификационные колонны, теплообменники, блоки механического деления и смешения потоков и др.). Требуется выбрать и соединить $\bar{N}$, $(\bar{N} \leqslant N)$ блоков таким образом, чтобы получить заданные количества целевых продуктов Y_i, $(i = \overline{1, m})$, обеспечив минимум приведенных затрат на построение схемы (или максимум прибыли). Ясно, что число возможных вариантов схем огромно и задача синтеза схемы сводится к сложной комбинаторной задаче определения наилучшей схемы из огромного числа возможных.

Введем понятие элементарного блока синтеза (ЭБС). Простейшим ЭБС является аппарат. Но ЭБС может состоять и из нескольких аппаратов (блоков) с заданной структурой потоков между ними; тогда задача синтеза схемы сводится к определению связей между ЭБС, внутри же каждого из них связи остаются неизменными. Примеры ЭБС будут приведены ниже. Ясно, что, чем больше аппаратов из N возможных заключены в ЭБС, тем меньше вариантов возможных схем, тем проще задача синтеза. В дальнейшем для простоты изложения часто будем говорить «блок», подразумевая элементарный блок синтеза.

Перейдем теперь к математической формулировке задачи синтеза. Введем следующие обозначения: α — вектор, компонентами которого являются все величины α_{ij}^{ks}; $\Psi(\alpha)$ — вектор, компонентами которого являются величины $\Psi^{(q, m)}$ [см. выражение (I,7)]. Условие (I, 7) в векторной записи примет вид

$$\Psi(\alpha) = 0 \qquad \text{(VI, 1)}$$

Будем исходить из предположения, что математические модели и критерий обладают следующими свойствами:

$$f^{(k)}(0, u^{(k)}) \equiv 0 \qquad F^{(k)}(0, \ 0, \ 0) = 0 \qquad \text{(VI, 2)}$$

Так как после задания всех значений параметров α и управлений u [см. выражение (1, 15)] значение критерия F подсчитывается с помощью уравнений (1, 1), (1, 2), F можно считать сложной функцией этих переменных $F = F(\alpha, u)$. Согласно формуле (I, 4а) структурные переменные α_{ki}^{sj} должны быть двоичными: $\alpha_{ki}^{sj} = 0$ или 1. Перепишем это условие в векторной форме

$$\alpha \text{ — целое} \qquad 0 \leqslant \alpha \leqslant 1 \qquad \text{(VI,3)}$$

Введем множество G^0 параметров α следующим образом:

$$G^0 = \{\alpha : \Psi(\alpha) = 0, \ \alpha — \text{целое}; \ 0 \leqslant \alpha \leqslant 1\} \qquad (VI, 4)$$

Поскольку каждому набору α, удовлетворяющему условиям (VI, 1), (VI,3), соответствует какая-либо схема, то множеству G^0 соответствуют все возможные варианты схем проведения процесса. Отсюда задача синтеза может быть записана в виде

$$\min_{\alpha \in G^0, \, u \in U} F(\alpha, u) \qquad (VI, 5)$$

В отличие от структурных переменных α, которые в данном случае являются дискретными, поисковые переменные будем называть технологическими, поскольку в качестве u используются обычно температуры, давления, расходы и конструктивные параметры. Как правило, это непрерывные переменные, хотя по условиям задачи некоторые из конструктивных параметров могут принимать только дискретные значения (например, в соответствии с ГОСТ, длины и диаметры труб теплообменников могут принимать только дискретный ряд значений).

Задача (VI, 5) относится к многоразмерным дискретно-непрерывным экстремальным задачам, методы решения которых еще не получили такого развития, как в случае непрерывных экстремальных задач. Остановимся на математических методах, которые используются для решения задач синтеза.

Методы дискретной математики [120] (целочисленное линейное программирование, методы поиска на дереве решений [121, с. 82], в том числе методы «ветвей и границ» [120, с. 213] и др.), использование которых обусловлено комбинаторным характером проблемы.

Локальные поисковые методы с непрерывными поисковыми переменными (см. гл. III, IV), получившие большое развитие в последние годы. Чтобы получить возможность их использования, желательно сводить задачу синтеза ХТС к экстремальной непрерывной задаче.

Методы поиска глобального экстремума [12, с. 491—535]. При оптимизации систем фиксированной структуры обычно используются локальные методы поиска, поскольку при этом либо известно хорошее начальное приближение, либо задача носит одноэкстремальный характер. Задачи же синтеза часто имеют многоэкстремальный характер, что существенно усложняет их решение [122] и приводит к необходимости применения методов глобальной оптимизации.

Метод погружения [123], часто используемый, когда первоначальную задачу заменяют более общей, имеющей развитые методы решения. При этом решение общей задачи дает решение и частной задачи. На этом принципе построен метод динамического программирования [124]. В данном случае эта идея реализуется следующим образом. Строится некоторая схема (назовем ее глобальной, или интегральной), которая охватит все частные схемы рассматриваемой задачи синтеза. Причем любую частную схему можно получить из глобальной, если некоторым параметрам (будем называть их структурными) давать определенные значения. Тогда

оптимизация глобальной схемы будет давать решение исходной задачи синтеза. Ясно, что глобальная схема должна строиться таким образом, чтобы можно было использовать развитые методы оптимизации, например нелинейное программирование, целочисленное линейное программирование и т. д.

Метод закрепления. Коротко он был рассмотрен в гл. V применительно к задаче оптимизации схем фиксированной структуры с непрерывными переменными, однако, он имеет силу и для решения задач синтеза. Действительно, рассмотрим схему, приведенную на рис. 22. Под отдельным блоком этой схемы будем понимать какую-либо подсхему. Закрепим все входные и выходные переменные блоков этой схемы. Поскольку при этом ликвидируется взаимное влияние блоков, улучшение структуры подсхемы относительно локального критерия не будет противоречить глобальному критерию.

Сложность задач синтеза требует максимального использования специфики при решении каждой отдельной задачи. В связи с этим получили большое развитие методы синтеза гомогенных схем, т. е. схем, состоящих из однородных аппаратов; примером могут служить теплообменные системы (ТС) [46, с. 257—308], системы разделения [125]. Рассмотрим методы синтеза как гетерогенных систем, т. е. схем состоящих из разнотипных аппаратов, так и гомогенных систем.

Двухуровневый подход

Задача синтеза (VI,5) записывается в виде следующей двухуровневой процедуры:

$$\min_{\alpha \in G^0} \min_{u \in U} F(\alpha, u) \qquad (VI, 6)$$

На первом уровне при фиксированных α проводится минимизация функции F по непрерывным переменным u [1, с. 304—306]. Поскольку задача оптимизации F при фиксированных α соответствует оптимизации некоторой схемы фиксированной структуры, на первом уровне могут использоваться хорошо разработанные методы оптимизации схем фиксированной структуры (см. гл. I—IV). Введем обозначения

$$F^*(\alpha) = \min_{u \in U} F(\alpha, u)$$

тогда на втором уровне должна решаться задача

$$\min_{\alpha \in G^0} F^*(\alpha) \qquad (VI, 7)$$

Таким образом, задача второго уровня сводится к минимизации нелинейной функции двоичных переменных. Одним из подходов к решению задач второго уровня (VI, 7) является полный перебор всех вариантов схем.

Полный перебор вариантов схем

Основная цель состоит в формализации процедуры перебора, которая обеспечивала бы получение всех допустимых вариантов схем. Для этого удобно использовать специальный вид ориентированного

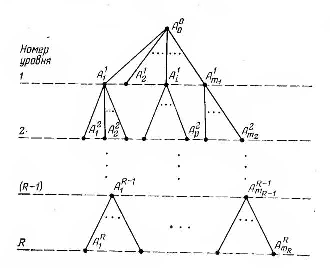


Рис. 34. Ориентированный граф-дерево.

графа, называемый д е р е в о м (рис. 34). Дерево характеризуется следующими свойствами: 1) имеет одну корневую вершину A_0^0; 2) все ребра направлены сверху вниз, поэтому стрелки не показаны; 3) каждая вершина имеет только одно входное ребро. Вершина называется в и с я ч е й, если она не имеет выходных ребер. Назовем k-тым у р о в н е м дерева совокупность вершин графа, обладающих таким свойством: число ребер, соединяющих вершину A_0^0 с любой вершиной k-го уровня, равно k (рис. 34). Будем считать, что в дереве имеется R уровней. Вершины j-го уровня обозначим через A_i^j (где j — номер уровня; i, $(i = \overline{1, m_j})$ — номер вершины данного уровня; m_j — число вершин на j-том уровне). Если ребро соединяет вершину A_i^j с вершиной A_p^{j+1}, то будем говорить, что A_i^j — п р е д ш е с т в у ю щ а я вершина, а A_p^{j+1} — п о т о м о к вершины A_i^j. Пусть p_i^j — число потомков вершины A_i^j, а M_i^j — множество номеров вершин $(j+1)$-го уровня, являющихся потомками вершины A_i^j. Будем называть д е р е в о м в а р и а н т о в дерево, с помощью которого могут быть перечислены все варианты схем. Рассмотрим три их типа.

1-й тип дерева вариантов. В этом случае вершина A_0^0 соответствует все множество вариантов схем, а вершине A_i^j некоторое его подмножество. Обозначим через $\overline{A}_i^j$ множество схем, соответствующих вершине A_i^j. В построении дерева вариантов основным является правило, согласно которому строятся потомки любой вершины. Такие правила могут иметь самый разный характер, однако они должны удовлетворять следующим трем условиям.

1. Множество схем, соответствующих любому потомку вершины A_i^j является подмножеством схем, соответствующих вершине A_i^j

$$\overline{A}_p^{j+1} \subset \overline{A}_i^j \qquad p \in M_i^j \tag{VI, 8}$$

2. Сумма множеств, соответствующих всем потомкам вершины A_i^j, равна множеству, соответствующему вершине A_i^j

$$\overline{A}_i^j = \bigcup_{p \in M_i^j} \overline{A}_p^{j+1} \tag{VI, 9}$$

3. Ни одна из схем не может одновременно входить в два подмножества, соответствующие двум потомкам вершины A_i^j

$$\overline{A}_p^{j+1} \cap \overline{A}_q^{j+1} = \varnothing \qquad p,\ q \in M_i^j \tag{VI, 10}$$

Ясно, что если вершине соответствует только одна схема, то она не имеет потомков и является висячей. Для примера приведем одно правило построения потомков. Согласно этому правилу, каждая вершина за исключением висячих имеет два потомка, одному из которых соответствует множество схем, в которых обязательно присутствует некоторый поток, а другому потомку — множество схем, в котором указанный поток отсутствует. Итак, каждому из двух ребер, выходящему из вершины A_i^j, соответствует одно условие (наличие потока в схеме или его отсутствие) построения схемы. Отсюда следует, что совокупность ребер, входящих в путь от вершины A_0^0 до любой висячей вершины A_i^j, полностью характеризует схему, соответствующую этой вершине. Ясно, что в данном случае число уровней будет равно числу возможных потоков. Если в схеме имеется N аппаратов, каждый из которых имеет только один выходной поток, то число возможных потоков будет равно $N\ (N-1)$ и, следовательно $R = N\ (N-1)$.

2-й тип дерева вариантов. В этом случае каждой вершине A_i^j ставится в соответствие только одна схема $\overline{A}_i^j$. Потокам A_p^{j+1}, $(p \in M_i^j)$ вершины A_i^j соответствуют схемы, которые получаются из схемы $\overline{A}_i^j$ добавлением одного аппарата в соответствии с каким-либо формализованным правилом.

Расширение схемы продолжают, пока не будут исчерпаны все аппараты, из которых должна быть построена схема. Естественно, что вершины последнего уровня будут висячими. Дерево вариантов 2-го типа было построено для задачи синтеза теплообменных систем [136].

3-й тип дерева вариантов. Ребра этого дерева соответствуют некоторым потокам. Для простоты изложения будем говорить, что в данную вершину входит (или из нее выходит) ребро, соответствующее этому потоку. Граф будет иметь вершины трех типов. Каждой вершине первого типа будет соответствовать какой-либо один аппарат. Если поток, выходящий из вершины 1-го типа, не требует дальнейшего преобразования, то ребра, выходящие из такой вершины,

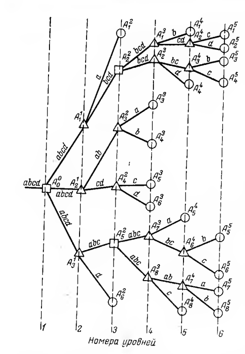


Рис. 35. Дерево вариантов схем разделения четырехкомпонентной смеси:
вершины: △ — 1-го типа; □ — 2-го типа; ○ — 3-го типа.

попадают либо в висячую вершину (относится к 3-му типу), либо в вершины 2-го типа. В этих вершинах надо принять следующее решение: в какой из еще не использованных аппаратов подать поток, выходящий из вершины 1-го типа. Здесь путь от вершины A_0^0 до промежуточной вершины A_i^j будет соответствовать части некоторой схемы. Путь от A_0^0 до некоторых висячих вершин будет соответствовать полной схеме.

Приведем пример дерева этого типа. На рис. 35 показано дерево вариантов всевозможных схем, состоящих из ректификационных колонн, предназначенных для разделения четырехкомпонентной смеси (*abcd*), где *a*, *b*, *c*, *d* — некоторые компоненты; над каждым ребром вписана соответствующая смесь. К вершинам 1-го типа принадлежат A_1^1, A_2^1, A_3^1, A_3^2 и др. Вершине A_1^1 соответствует ректификационная колонна, в которой смесь (*abcd*) делится на потоки *a* и (*bcd*), эту операцию будем обозначать через $(a|bcd)$. К вершинам 2-го типа принадлежат A_0^0, A_2^2, A_5^2. Из A_0^0 ребра попадают в вершины A_1^1, A_2^1, A_3^1, соответствующие ректификационным колоннам, в которых смесь (*abcd*) делится соответственно, следующим образом: $(a|bcd)$, $(ab|cd)$, $(abc|d)$. К вершинам 3-го типа принадлежат A_1^2, A_6^2, A_1^4, A_1^5 и др. Путь $A_0^0A_1^5$ соответствует схеме разделения, состоящей из 3-х ректификационных колонн, в которых разделение смеси проводится в следующей последовательности: $(a|bcd)$, $(b|cd)$, $(c|d)$.

После того, как дерево вариантов одного из трех типов построено, процедуру перебора вариантов легко формализовать. Действительно, согласно способу построения любого из трех типов деревьев совокупность всех (или некоторых) висячих вершин дерева соответствует всем возможным вариантам схемы. Отсюда становится ясным принцип перебора — необходимо просмотреть все висячие вершины дерева. Для этого может быть использован так называемый перебор «в глубину» [121, с. 103], при котором последовательно просматриваются пути, связывающие вершину A_0^0 со всеми висячими вершинами. При использовании дерева 2-го типа, когда каж-

дой вершине соответствует какая-либо схема, возможен также перебор «в ширину» [121, с. 102], при котором последовательно просматриваются сначала вершина A_0^0, потом все вершины, начиная с первого уровня и кончая последним. В данном случае возможен также перебор «на глубину k», когда просматриваются перебором «в ширину» все вершины от нулевого до k-го уровня. Из этих вершин выбирается та, которой соответствует наилучшая схема, и из нее производится перебор «в ширину» вершин следующих k уровней и т. д.

Несмотря на то, что процесс перебора поддается полной формализации, его нельзя применять в любом случае, поскольку число вариантов схем может быть астрономическим. Так например, если число аппаратов $N = 10$ и все аппараты имеют по одному выходу, то число P структурных параметров α будет равно $N^2 = 100$. Каждый параметр может принимать значение 0 или 1, поэтому с учетом ограничений (I, 7) общее число вариантов будет равно 100! Отсюда ясно, что полный перебор неприемлем даже для такой сравнительно небольшой задачи. Необходимы средства, которые существенно ускоряли бы процедуру перебора. Рассмотрим два из них: метод ветвей и границ [127, с. 299] и метод отсечения неперспективных вариантов.

Метод „ветвей и границ"

Процедуру метода «ветвей и границ» удобно представить в виде дерева решений (рис. 36). В отличие от дерева вариантов (рис. 34) вершины дерева решений будут иметь либо два, либо один, либо ни одного потомка. Так же, как и в дереве вариантов, вершине A_0^0 соответствует множество $\overline{A}_0^0$ всех схем, которые можно составить из N блоков, а вершине A_i^j — некоторые подмножества схем $\overline{A}_i^j$ $(\overline{A}_i^j \subset \overline{A}_0^0)$. Внутри каждого подмножества $\overline{A}_i^j$ введем свою нумерацию схем, через a_{ik}^j обозначим k-тую схему множества A_i^j. Пусть $F^*(a_{ik}^j)$ — оптимальное значение критерия (I,15) для схемы a_{ik}^j

$$F^*(a_{ik}^j) = \min_{u \in U} F(\alpha, u) \qquad \text{(VI, 11)}$$

Нижней оценкой ρ_i^j множества $\overline{A}_i^j$ (вершины A_i^j) называется число, обладающее свойством

$$\rho_i^j \leqslant F_i^{j*} = \min_k F^*(a_{ik}^j) \qquad \text{(VI, 12)}$$

Если A_p^{j+1} является потомком A_i^j, то $\overline{A}_p^{j+1} \in \overline{A}_i^j$; отсюда вытекает очевидное неравенство

$$\rho_p^{j+1} \geqslant \rho_i^j \qquad (\overline{A}_p^{j+1} \subset \overline{A}_i^j) \qquad \text{(VI, 13)}$$

Величина F_i^{j*} является оптимальным значением критерия для наилучшей схемы из множества $\overline{A}_i^j$. Будем называть эту величину

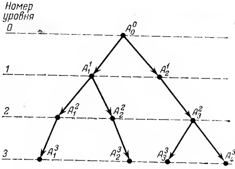


Рис. 36. Дерево решений.

нижней гранью множества $\overline{A}_i^j$. В процедуре метода ветвей и границ j-му шагу соответствует j-тый уровень дерева решений (см. рис. 36). На каждом шаге этой процедуры проделываются следующие операции.

1. Находятся нижние оценки ρ_i^j всех множеств A_i^j, $(i = \overline{1, m_j})$;

2. Определяется множество $\overline{A}_{p_j}^j$, для которого нижняя оценка будет наименьшей, т. е. существует неравенство

$$\rho_{p_j}^j \leqslant \rho_i^j \qquad i = \overline{1,\ m_j} \tag{VI, 14}$$

Это множество будем называть «оптимальным», а соответствующую вершину — «оптимальной». Заметим, что если бы в соотношении (VI, 12) выполнялось точное равенство, то оптимальная схема обязательно принадлежала бы «оптимальному» множеству. В этом случае дальнейшему рассмотрению необходимо было бы подвергнуть только множество $\overline{A}_{p_j}^j$, отбросив все остальные, как заведомо не содержащие оптимальной схемы. В случае же неравенства не исключена возможность того, что в действительности оптимальная схема не принадлежит множеству $\overline{A}_{p_j}^j$ (чем ближе величина ρ_i^j к F_i^{j*}, тем менее вероятна такая ситуация), что заставляет анализировать и «неоптимальные» множества $\overline{A}_i^j$, $(i \neq p_j)$. Разобьем неоптимальные вершины на перспективные и неперспективные вершины. Вершину A_i^j, $(i \neq p_j)$ будем называть неперспективной, если множество $\overline{A}_i^j$ заведомо не содержит оптимальной схемы. Если множество $\overline{A}_i^j$, вообще говоря, может содержать оптимальную схему, то вершину A_i^j, $(i \neq p_j)$ будем называть перспективной.

3. Проверяется критерий окончания процедуры. При его выполнении оптимальная схема найдена и поиск прекращается. В противном случае осуществляется переход к следующей операции.

4. Оптимальное множество $\overline{A}_{p_j}^j$ разбивается на два подмножества. Соответствующие этим двум подмножествам вершины будут принадлежать $(j + 1)$-му уровню.

5. Среди неоптимальных вершин определяются перспективные и неперспективные. Неперспективные исключаются из рассмотрения, а для каждой перспективной вершины A_i^j, $(i \neq p_i)$ образуется один потомок A_q^{j+1}, которому соответствует то же множество схем, что и вершине A_i^j, $(\overline{A}_i^j = \overline{A}_q^{j+1},\ q \neq p_j)$.

Так, на рис. 37 множество $\overline{A}_3^2$ предполагается оптимальным и вершина A_3^2 имеет два потомка, а неоптимальные вершины A_1^2, A_2^2 предполагаются перспективными, они имеют по одному потомку. Итак, для выполнения процедуры метода «ветвей и границ» необходимо иметь следующие алгоритмы:

1) способ определения нижних оценок;

2) правило разбиения «оптимального» множества (правило ветвления);

3) критерий окончания процедуры;

4) способ определения перспективных и неперспективных среди «неоптимальных» вершин.

Опишем теперь общие подходы, которые могут быть положены в основу построения этих алгоритмов.

Определение нижней оценки. Пусть имеется задача

$$\min f(x) \qquad x \in G \tag{VI, 15}$$

и требуется найти нижнюю оценку функции $f(x)$ на множестве (области) G. Рассмотрим три способа определения нижней оценки.

1 способ — переход к большей области. Пусть имеется множество $\overline{G}$, такое, что $G \subset \overline{G}$, тогда выполняется очевидное соотношение

$$\min_{x \in \overline{G}} f(x) \leqslant \min_{x \in G} f(x) \qquad G \subset \overline{G} \tag{VI, 16}$$

Отсюда следует, что в качестве нижней оценки функции $f(x)$ на множестве G может быть взята величина

$$\rho = \min_{x \in \overline{G}} f(x) \tag{VI, 17}$$

При решении задачи (VI,17) необходимо искать глобальный минимум. Область $\overline{G}$ выбирается так, чтобы решение задачи (VI,17) было в каком-то смысле проще решения задачи (V,15). Так например, если G — невыпуклая область, то задача (VI,15) может иметь несколько минимумов и для ее решения необходимо использовать сложные нелокальные методы спуска. Если в качестве $\overline{G}$ взять выпуклую область, то для решения задачи (VI,17) можно воспользоваться хорошо разработанными методами локального спуска.

2 способ — переход от целочисленных переменных к непрерывным. Пусть, например, область G представляет собой совокупность точек с целочисленными координатами, лежащими в некотором гиперкубе

$$G = \{x : a_i \leqslant x_i \leqslant b_i \qquad x_i \text{ — целые}\} \tag{VI, 18}$$

При этом различают два случая. В первом из них функция $f(x)$ определена и может быть рассчитана во всех точках гиперкуба; в точках с нецелочисленными координатами она либо не имеет физического смысла, либо же не может реализоваться в силу каких-либо условий (например, дискретность значений длин и диаметров труб теплообменников, определяемых ГОСТом). Здесь мы имеем дело с частным случаем 1-го способа, поскольку от минимизации

в области G мы переходим к минимизации в области $\overline{G}$, определяемой условиями

$$\overline{G} = \{x: \quad a_i \leqslant x_i \leqslant b_i\} \tag{VI, 19}$$

Ясно, что $G \subset \overline{G}$ и величина ρ из выражения (VI,17) будет нижней оценкой функции $f(x)$. Более сложным будет случай, когда функция $f(x)$ не определена в точках с нецелочисленными координатами; здесь необходимо построить такую функцию $\bar{f}(x)$, которая совпадает с значениями функции $f(x)$ в точках с целочисленными координатами и определена в остальных точках гиперкуба (VI,19):

$$\bar{f}(x) = \begin{cases} f(x), \text{ если } x \in G \\ \text{определена, если } x \in \overline{G}/G \end{cases} \tag{VI, 20}$$

Ясно, что в качестве нижней оценки в этом случае может быть взята величина

$$\rho = \min_x \bar{f}(x) \qquad x \in \overline{G} \tag{VI, 21}$$

3 способ — переход от конечномерных задач к бесконечномерным. Известные задачи определения оптимального профиля температур, давлений, нагрузок реактора [128, с. 46] по существу дают нижнюю (верхнюю) оценку для всевозможных стратегий изменения указанных параметров по длине реактора.

В заключение отметим, что задача разработки алгоритма нахождения нижних оценок множеств является центральной и наиболее сложной из возникающих при использовании метода ветвей и границ. К сожалению, именно эта часть процедуры является нестандартной, и ее приходится разрабатывать для каждой задачи в отдельности. От того, насколько эффективной будет эта процедура, во многом зависит успех применения метода ветвей и границ. Причем при разработке этого алгоритма приходится решать компромиссную задачу. С одной стороны, чем точнее алгоритм, т. е. чем ближе величина ρ_i^j к точке нижней грани F_i^{j*}, тем больше вероятность нахождения оптимальной схемы в множестве $\overline{A}_{p_j}^j$ и тем меньше будет ложных ветвлений. С другой стороны, более точный алгоритм обычно требует больше времени для определения ρ_i^j.

Окончание процедуры поиска оптимальной ХТС определяется следующими условиями.

1. Среди схем множества $A_{p_j}^j$ найдена схема номер $\bar{l}$, для которой выполняется условие

$$\rho_{p_j}^j = F_{p_j}^{j*} \qquad F_{p_j}^{j*} = F^*\left(a_{i\bar{l}}^j\right) \tag{VI, 22}$$

2. Среди схем множества $A_{p_j}^j$ найдена схема номер $\bar{\bar{l}}$, для которой выполняется условие

$$\left|\rho_{p_j}^j - F^*\left(a_{p_j\bar{\bar{l}}}^j\right)\right| \leqslant \varepsilon \tag{VI, 23}$$

где ε — достаточно малая величина.

Действительно, пусть выполняется условие (VI,22), тогда l-ая схема является наилучшей среди схем множества $\overline{A}^j_{p_j}$. С другой стороны, в соответствии с неравенством (VI,14) ни в одном из множеств A^j_i, $(i \neq p_j)$ не найдется схемы, лучшей чем a^j_{il}. В частности условие (VI,22) выполняется, если в результате предыдущих делений множество $\overline{A}^j_{p_j}$ оказалось состоящим только из одной схемы. В случае выполнения условия (VI,23) можно показать, что нельзя найти схему, для которой выполнялось бы условие

$$F\left(a^j_{qk}\right) < F\left(a^j_{p_j\bar{l}}\right) - \varepsilon \qquad q \neq p_j$$

Отсюда следует, что в качестве оптимальной с достаточной точностью может быть принята схема $a^j_{p_j\bar{l}}$.

Правило разбиения «оптимального» множества (правило ветвления). Поскольку в оптимальной вершине соответствующее множество делится на два множества, то иногда говорят о ветвлении в «оптимальной» вершине, а правило разбиения называют правилом ветвления. Для формализации процедуры ветвления введем вектор параметров δ, компоненты которого δ_i, $(i = \overline{1, p})$ могут принимать значения 0 или 1. Задание фиксированных значений 0 или 1 для одного или нескольких из этих параметров будет выделять среди множества схем $\overline{A}^0_0$ некоторое подмножество. В отличие от структурных параметров параметры δ_i будем называть структурными параметрами 2-го рода, их смысл будет конкретизироваться в каждом отдельном случае.

Структурный параметр 2-го рода будет характеризовать процедуру ветвления в «оптимальной» вершине. Так, множество схем, соответствующее потомку A^{j+1}_q оптимальной вершины $A^j_{p_j}$ будет характеризоваться тем, что для него один из параметров (например, δ_s) будет принимать значение 1; множество схем, соответствующее другому потомку A^{j+1}_r, будет характеризоваться значением $\delta_s = 0$. Параметр δ_s будем называть параметром ветвления вершины $A^j_{p_j}$ (множество $\overline{A}^j_{p_j}$). Кроме того, будем говорить, что ребрам $A^j_{p_j}$, A^{j+1}_q и $A^j_{p_j}$, A^{j+1}_r соответствуют значения параметра δ_s, равные 1 и 0, соответственно. Итак: каждой «оптимальной» вершине будет соответствовать свой параметр ветвления. Множество схем $\overline{A}^j_i$ будет характеризоваться некоторым множеством $D^j_i = \overline{D}^j_i \cup D^j_i$ параметров δ_q. Это множество будет формироваться следующим образом. Рассмотрим путь $A^0_0 A^j_i$, связывающий вершины A^0_0 и A^j_i на дереве решений (см. рис. 36). Занесем в множество $\overline{D}^j_i$ значения параметров ветвления, соответствующих всем ребрам, принадлежащим этому пути и выходящим из «оптимальных» вершин этого пути. В общем случае $\overline{D}^j_i$ будет иметь вид

$$\overline{D}^j_i = \{\delta_i : \delta_{p_1} = 1, \ldots, \delta_{p_{ij}} = 1;\ \delta_{q_1} = 0, \ldots, \delta_{q_{ij}} = 0\} \qquad \text{(VI, 24)}$$

Множество $\overline{\overline{D}}_l^j$ будет содержать все δ_l, не вошедшие в $\overline{D}_l^j$

$$\overline{\overline{D}}_l^j = \{\delta_l : \delta_l \notin \overline{D}_l^j\}$$

Ясно, что $\overline{D}_0^0 = \varnothing$. Пусть параметром ветвления множества $\overline{A}_0^0$ был параметр δ_l, причем схемы множества $\overline{A}_1^1$ характеризуются тем, что для них $\delta_l = 1$, а для схем множества $\overline{A}_2^1$ $\delta_l = 0$, тогда $\overline{D}_1^1 = \{\delta_l = 1\}$, $\overline{D}_2^1 = \{\delta_l = 0\}$. Пусть на первом уровне вершина A_1^1 — «оптимальная», а ее параметром ветвления является параметр δ_r. Пусть схемы множества $\overline{A}_1^2$ характеризуются тем, что для них $\delta_r = 1$, а для схем множества $\overline{A}_2^2$ $\delta_r = 0$. Тогда $\overline{D}_1^2 = \{\delta_l = 1,\ \delta_r = 1\}$, $\overline{D}_2^2 = \{\delta_l = 1,\ \delta_r = 0\}$. Поскольку множества $\overline{A}_3^2$ и $\overline{A}_2^1$ совпадают, то $\overline{D}_3^2 = \overline{D}_2^1$. В дальнейшем нам понадобится вспомогательный вектор $\Delta_l^j = \{\bar{\delta}_1, \ldots, \bar{\delta}_p\}$, компонентами которого будут структурные параметры $\bar{\delta}$

$$\bar{\delta}_k = \begin{cases} \delta_k, & \text{если } \delta_k \in \overline{D}_l^j \\ 1, & \text{если } \delta_k \in \overline{\overline{D}}_l^j \end{cases} \qquad \text{(VI, 25)}$$

Определение неперспективных неоптимальных вершин. Для проверки критерия окончания процедуры поиска [по условию (VI, 23)] в каждой «оптимальной» вершине всегда будем брать некоторую схему $a_{pj\bar{l}}^j$, для которой будем рассчитывать оптимальное значение критерия $F^*(a_{pj\bar{l}}^j)$. В каждом отдельном случае правило выбора этой схемы будет различным. Из определения нижней оценки вытекает, что если выполняется неравенство

$$\rho_l^j > F^*\left(a_{pj\bar{l}}^j\right)$$

то все схемы множества $\overline{A}_l^j$ хуже, чем схема a_{pjl}^j. Отсюда следует, что все вершины, в которых удовлетворяется это условие — неперспективны и должны быть исключены из дальнейшего анализа.

Метод отсечения неперспективных вариантов

Метод отсечения неперспективных вариантов применяется к дереву вариантов 3-го типа. Для иллюстрации рассмотрим пример дерева вариантов схем разделения (см. рис. 35). Обозначим через F_l^{j*} затраты на часть схемы, соответствующую пути $A_0^0 A_l^j$. Если вершина A_p^q принадлежит пути $A_0^0 A_l^j$, то подсхема, соответствующая пути $A_0^0 A_p^q$, является частью подсхемы, соответствующей пути $A_0^0 A_l^j$. Поскольку на часть схемы всегда приходятся меньшие затраты, чем на всю схему, справедливо соотношение

$$F_p^{q*} < F_l^{j*} \qquad A_p^q \subset A_0^0 A_l^j$$

На основе этого неравенства может быть построено правило отсечения неперспективных вариантов. Действительно, пусть из-

вестны затраты F_a^{b*} на полную схему, соответствующую некоторой висячей вершине A_a^b. Пусть в некоторой вершине A_p^q соблюдается неравенство

$$F_p^{q*} > F_a^{b*}$$

В этом случае нет смысла рассматривать всех потомков вершины A_p^q, а также потомки ее потомков и т. д., поскольку соответствующие схемы будут хуже, чем схема, соответствующая висячей вершине A_a^b. Поясним это правило на примере дерева вариантов схем разделения (см. рис. 35). Пусть для просмотра вариантов используется перебор «в глубину» и определен первый путь $A_0^0 A_1^5$, которому соответствует схема разделения с затратами F_1^{5*}. После этого двигаясь из вершины A_0^0, следует проверить путь, проходящий через вершину A_2^3. Если окажется, что $F_2^{3*} > F_1^{5*}$, то вершину A_3^4 можно не рассматривать, поскольку по ее достижении стоимость схемы может только возрасти [129]. Общая схема применения этого правила будет следующая. Во время перебора «в глубину» в качестве верхней границы берутся наименьшие величины из затрат, соответствующих просмотренным схемам. Если во время перебора в некоторой промежуточной вершине величина затрат окажется выше верхней границы, то продолжать просмотр потомков этой вершины нет необходимости.

Остановимся теперь на способе вычисления величины F_p^{q*}, воспользовавшись схемой, приведенной на рис. 35. Будем исходить из предположения, что каждая колонна обладает высокой разделительной способностью и во всех колоннах давление одинаково (изобарическая схема разделения). Рассмотрим вначале отдельную колонну, пусть это будет ректификационная колонна соответствующая вершине A_3^1 (см. рис. 35). При достаточной степени чистоты дистиллята и кубового продукта с некоторым приближением можно считать, что дистиллят состоит только из компонент *abc* (причем величины этих компонент будут равны их значениям на входе в колонну), а кубовый продукт — из компоненты *d*. Температура дистиллята будет равна температуре точки росы или кипения смеси *abc*, а кубового продукта — температуре кипения вещества *d*. Поэтому, как бы мы ни варьировали параметры колонны, ее выходные переменные будут постоянны и оптимизация никак не повлияет на режим работы остальных колонн. Таким образом, при оптимизации изобарической схемы ректификационных колонн каждую колонну можно оптимизировать отдельно. Отсюда возможен следующий простой способ вычисления F_l^{j*}. Пусть F_l^{j*} подсчитано и необходимо вычислить $F_p^{(j+1)*}$, где A_p^{j+1} — потомок вершины A_l^j. Для этого надо провести оптимизацию колонны, соответствующей вершине A_p^{j+1}, и величину полученных затрат для этой колонны прибавить к F_l^{j*}.

Метод структурных параметров

Будем исходить из предположения, что в задаче (VI, 5) проводится одновременная минимизация по дискретным переменным α и непрерывным переменным u (одноуровневый подход). Рассмотрим метод решения этой задачи, основанный на замене целочисленных переменных непрерывными [130], т. е. тот, который рекомендовался для получения нижних оценок в процедуре метода «ветвей и границ». Итак, будем считать структурные параметры непрерывными, удовлетворяющими условиям

$$0 \leqslant \alpha \leqslant 1 \tag{VI, 26}$$

Тогда полученная задача может быть записана в виде

$$\min_{\alpha \in \overline{G}_0^0,\ u \in U} F(\alpha, u) \tag{VI, 27}$$

где

$$\overline{G}_0^0 = \{\alpha : \Psi(\alpha) = 0, \qquad 0 \leqslant \alpha \leqslant 1\} \tag{VI, 28}$$

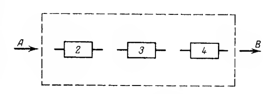
Дадим задаче (VI, 27) схемную интерпретацию. Будем считать' что стоимость узлов смешения и деления потоков намного меньше стоимости остальных аппаратов и ею можно пренебречь (в большинстве случаев это так). На каждом входном потоке схемы поставим мнимый входной блок; через каждый из таких блоков потоки веществ проходят без всякого преобразования. Новый ЭБС построим следующим образом. На каждом выходном потоке каждого оставшегося и мнимого блока поставим делитель, который будет распределять поток между входами всех блоков (за исключением мнимых), а на входном потоке каждого блока предусмотрим блок смешения, куда будут поступать доли потоков из всех блоков схемы. Совокупность блока, всех его выходных делителей потоков и входных смесителей потоков и будет новым ЭБС. Теперь становится ясным смысл введения мнимых входных блоков — он состоит в обеспечении возможности разветвления входных потоков схемы. Обозначим через α_{ki}^{sj} величину, удовлетворяющую условию (VI, 26) и показывающую, какая часть j-го выходного потока i-го блока подается на s-тый вход k-го блока.

Из уравнений материального баланса делителя потоков, стоящего на q-том выходном потоке m-го блока, и смесителя потоков, стоящего на s-том входном потоке k-го блока, легко получить соотношения (I, 7), (I, 4) соответственно. Таким образом, эти соотношения соблюдаются, когда структурные переменные являются как двоичными, так и непрерывными, удовлетворяющими условиям (VI, 26). Поскольку при различных значениях параметров α_{ki}^{sj} из этой схемы могут быть получены все возможные варианты, построенная схема будет глобальной.

Рассмотрим в виде примера задачу построения схемы, состоящей из трех аппаратов (рис. 37) (реактора идеального смешения, реактора идеального вытеснения, сепаратора и необходимого числа смесителей и делителей потоков. Пусть в обоих реакторах протекает

Рис. 37. Совокупность блоков, из которых должна быть построена схема получения продукта B из вещества A:

1 — мнимый блок (на схеме не показан); 2 — реактор идеального смешения; 3 — реактор идеального вытеснения; 4 — сепаратор. Пунктиром ограничена совокупность блоков, введение связей между которыми определяет некоторую схему проведения процесса.



необратимая изотермическая реакция первого порядка $A \xrightarrow{k_1} B$. Вещество B является целевым. В сепараторе поток вещества A отделяется от потока вещества B. Возможны различные способы организации схемы процесса. Исходный поток A можно подать вначале в реактор идеального смешения, а потом в реактор идеального вытеснения, либо в обратном порядке, либо его можно распределить между этими реакторами; можно вообще не использовать один из реакторов; поток между реакторами можно разделить, и часть его направить в сепаратор; непрореагировавшее вещество после сепаратора можно подать обратно на вход одного из двух реакторов; можно организовать частичный рецикл потока до его подачи в сепаратор и т. д.

Построим глобальную схему для этой задачи синтеза. На входных потоках блоков *2—4* поставим смесители *5—7* (рис. 38), а на выходных потоках — делители потоков *9—11*. Введем мнимый блок *1*, на выходном потоке которого поставим делитель *8*, цель которого — разделить исходный поток вещества A на два потока, подаваемых в блоки *2*, *3*. Все возможные варианты схем данной задачи содержатся в глобальной схеме, приведенной на рис. 38. Здесь через $x_1^{(k)}$, $x_2^{(k)}$ обозначены величины потока веществ A, B на входе k-го блока, а через $y_1^{(k)}$, $y_2^{(k)}$ потоки веществ A, B на выходе k-го блока. Поисковыми переменными будут структурные параметры α_{ij}, а также объемы реакторов. Элементарными блоками синтеза здесь являются совокупность блока i ($i = 2, 3, 4$), смесителя $i + 3$ и делителя $i + 7$, а также блок *1* и делитель *8*.

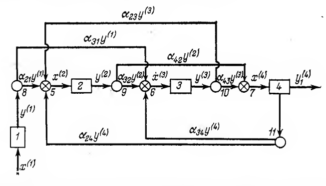


Рис. 38. Глобальная схема получения продукта B из вещества A (см. рис. 37):

1 — мнимый блок; 2 — реактор идеального смешения; 3 — реактор идеального вытеснения; 4 — сепаратор; 5—7 — смесители; 8—11 — делители потоков.

Задача оптимизации глобальной схемы будет иметь вид (VI, 27). Поскольку в этом случае все переменные являются непрерывными, для решения могут быть использованы хорошо разработанные численные методы нелинейного программирования (см. гл. III, IV). Ясно, что в результате решения могут быть получены нецелочисленные значения α_{ki}^{sj}, принимающие любые значения в интервале (VI, 26). Если условия задачи допускают любые значения структурных параметров в интервале (VI, 26), то полученный результат будет решением первоначальной задачи (VI, 5). При этом, если какие-либо структурные параметры α_{ki}^{sj} при $k = k_1, \ldots, k_p$; $s = \overline{1, N_k}$ примут нецелые значения, то на j-том выходе i-го блока необходимо поставить делитель потока, а на входных потоках блоков $k_1, \ldots, k_p$ смесители. В дальнейшем этот метод будем называть методом структурных параметров (МСП). Рассмотренный подход выглядит очень заманчивым, поскольку позволяет сводить многомерную комбинаторную задачу к задаче нелинейного программирования. Особенности этой задачи состоят в следующем:

1. Задача синтеза ХТС сводится к задаче нелинейного программирования (VI, 27) большой размерности. Если ХТС должна быть синтезирована из N блоков, то число P структурных параметров α_{ij} будет равно $N^2 + kN$ (где k — число входных потоков в схему). Так, даже для сравнительно небольшой задачи, когда $N = 10$, $k = 1$, число структурных параметров будет равно 110. В связи с этим для эффективности решения задач синтеза ХТС очень важно развитие методов решения поисковых задач большой размерности (см. гл. V).

2. В задаче нелинейного программирования (VI, 27) на все структурные переменные наложены простейшие ограничения, имеющие вид неравенств типа «больше—меньше», а также линейные ограничения типа (I, 7). Использование метода штрафа и метода уровней для учета этих ограничений может существенно ухудшить характеристики поиска, поэтому при решении задач синтеза важную роль должны играть методы поиска с непосредственным учетом линейных ограничений.

3. Использование МСП часто сводит задачу синтеза к многоэкстремальной задаче, что существенно усложняет ее решение, поскольку требует применения методов глобальной оптимизации [122].

4. В ряде случаев МСП в принципе не может преодолеть комбинаторный характер задачи синтеза ХТС.

Обобщение метода структурных параметров

Рассмотрим ряд важных для практических приложений случаев, когда МСП не позволяет полностью преодолеть комбинаторный характер задачи синтеза [132—134].

Случай 1. В результате решения задачи (VI, 27) оптимальные значения α^* структурных параметров могут принять нецелые значения. Имеется, однако, ряд случаев, когда требуется, чтобы прин-

ципиально выполнялось требование целочисленности параметров α. Эти требования могут возникнуть, если по тем или иным причинам нельзя смешивать некоторые потоки, например, вследствие взрывоопасности, или при применении МСП в задаче синтеза теплообменных систем. Пусть решение задачи (VI, 27) дало нецелые значения α^*, что может быть в частности следствием неполной сходимости к решению. В работе [130] предлагается использовать инженерную интуицию для решения вопроса о том, какие значения (0 или 1) должны принять структурные параметры α_{ij}. Однако использование этой рекомендации не всегда приведет к правильным результатам. Простое же округление полученных значений α^*_{ij} до ближайших целых чисел может привести к неправильным результатам [120, с. 17]. Возникает сложная проблема нахождения целых значений α_{ij} по полученным нецелым значениям α^*_{ij}. Решение этой задачи требует перебора различных вариантов значений структурных параметров α_{ij}.

Случай 2. Наличие фазовых ограничений (I, 11) также существенно затрудняет применение МСП. В работе [131] было показано, что применение этого метода в задаче синтеза теплообменных систем может дать неоптимальное решение при наличии ограничений на входные и выходные температуры теплообменников. Такая же ситуация может возникнуть и при его применении для синтеза более общих схем в случае наличия ограничений на фазовые переменные. Действительно, подставим в уравнение (I, 811) выражение для $x^{(k)}$ из формулы (I, 6)

$$\varphi^{(k)}\left(\sum_{j=1}^{N}\alpha_{kj}z^{(j)}\right)\leqslant a^{(k)} \qquad \text{(VI, 29)}$$

Из этого соотношения видно, что наличие фазовых ограничений в k-том блоке по существу накладывает ограничение на работу всех блоков схемы, поскольку в левую часть неравенства (VI, 29) входят выходные переменные всех блоков схемы. Пусть задача синтеза ХТС решена с помощью МСП и получено $u^{(k)} = 0$. Несмотря на то, что в этом случае $z^{(k)} = 0$, k-тый блок будет формально влиять на остальную схему вследствие необходимости соблюдения неравенства (VI, 29). Следовательно, ответ на вопрос о включении k-го блока в схему может быть дан только в результате решения двух задач синтеза, в одной из которых k-тый блок заранее учитывается в задаче, а во второй не учитывается. Поскольку ограничения на входные переменные могут существовать в нескольких блоках, возникает комбинаторная проблема выбора оптимальной комбинации из всех возможных вариантов включения или невключения в схему блоков, имеющих ограничения на входные переменные. Простой перебор может привести к очень большим величинам времени счета.

Случай 3. Возможна ситуация, при которой величины $F^{(k)}$ в критерии (I, 15) имеют постоянные составляющие, не зависящие от переменных $x^{(k)}$, $z^{(k)}$, $u^{(k)}$. Такое положение может создаться,

если, например, в $F^{(k)}$ учитывать затраты на системы автоматического управления, измерительные приборы и т. д. Известно, что с одной стороны эти затраты в настоящее время составляют существенную часть стоимости основных аппаратов, а с другой, мало зависят от конструктивных и технологических параметров этих аппаратов. Поэтому можно считать, что в данном случае критерий оптимизации имеет вид

$$F = \sum_{k=1}^{N} \{F^{(k)}(x^{(k)}, u^{(k)}, z^{(k)}) + P^{(k)}\} \qquad \text{(VI, 30)}$$

где $P^{(k)} \geqslant 0$ не зависит от $x^{(k)}$, $z^{(k)}$, $u^{(k)}$, $F^{(k)}(0, 0, 0) = 0$.

Аналогично предыдущему можно показать, что наличие в критерии частей $P^{(k)}$ также приводит к комбинаторной проблеме перебора всех возможных вариантов включения или невключения блоков в схему

Случай 4. Пусть в схеме имеются две (или несколько) группы потоков, которые должны удовлетворять следующему требованию. Внутри одной группы потоки могут смешиваться, потоки же из разных групп смешиваться не должны. Такая ситуация может возникнуть, например, из требований взрывобезопасности, или вследствие того, что потоки находятся в разных фазах. Естественно, что решение α^* задачи (VI, 27), вообще говоря, не будет удовлетворять этому требованию, вследствие чего также возникает комбинаторная проблема выбора оптимальной комбинации из всех возможных вариантов подачи в каждый блок потоков только одной группы. Возможен и смешанный случай, когда одна задача синтеза ХТС будет комбинацией четырех случаев рассмотренных выше. Подход, который будет рассмотрен для каждого из четырех случаев в отдельности, легко может быть обобщен и на указанный смешанный случай.

Для простоты изложения будем считать, что каждый блок имеет один входной и один выходной поток и воспользуемся обозначениями (I, 5). Излагаемый подход будет комбинацией метода «ветвей и границ» и МСП, который будет использоваться для получения нижних оценок, поскольку переход от дискретных переменных к непрерывным позволяет применять численные методы нелинейного программирования. Общая схема метода «ветвей и границ» для всех четырех случаев совпадает с описанной выше. Для каждого отдельного случая опишем правило ветвления, правило определения нижней оценки и правило окончания процедуры [132—134].

Синтез ХТС при условии целочисленности структурных параметров

Будем исходить из предположения, что критерий имеет вид (I, 15), а фазовые ограничения (I, 11) отсутствуют. Для простоты изложения примем, что все структурные параметры α_{ij} целочислены и имеется только один входной поток, подаваемый только в один блок.

Правило ветвления. В качестве структурных параметров 2-го рода выберем сами структурные параметры α_{ij}. Между этими параметрами будет существовать следующее взаимно однозначное соответствие:

$$\delta_k = \alpha_{ij} \qquad k = iN + j \qquad 1 \leqslant k \leqslant N^2 \qquad \text{(VI, 31)}$$

Пусть вершине $A^j_{p_j}$ соответствует параметр ветвления δ_p. Тогда множество схем $\overline{A}^j_{p_j}$ разбивается на два подмножества, соответствующие двум потомкам вершины $A^j_{p_j}$; в схемах одного из них присутствует поток, соответствующий значению параметра $\delta_p = 1$, в схемах другого этот поток отсутствует ($\delta_p = 0$).

Для множества схем $\overline{A}^j_i$ характерно следующее: потоки, соответствующие параметрам $\delta_{p_1}, \ldots, \delta_{p_{ij}}$ [см. выражение (VI, 24)] присутствуют во всех схемах, а потоки, соответствующие параметрам $\delta_{q_1}, \ldots, \delta_{q_{ij}}$, отсутствуют. Введем множества G^j_i и $\overline{G}^j_i$:

$$G^j_i = \{\alpha : \psi(\alpha) = 0,\ 0 \leqslant \alpha \leqslant 1,\ \alpha - \text{целое},\ \alpha \in D^j_i\} \qquad \text{(VI, 32)}$$

$$\overline{G}^j_i = \{\alpha : \psi(\alpha) = 0,\ 0 \leqslant \alpha \leqslant 1,\ \alpha \in D^j_i\} \qquad \text{(VI, 33)}$$

где в соответствии с соотношением (VI, 31) в D^j_i вместо δ_i стоят параметры α_{ij}. Множество $\overline{G}^j_i$ отличается от множества G^j_i только тем, что условие целочисленности параметров α заменено условием непрерывности и принадлежности интервалу (0, 1). Поскольку $D^0_0 = \varnothing$, то G^0_0 и $\overline{G}^0_0$ определяются выражениями (VI, 4), (VI, 28).

Определение нижних оценок. Задача определения F^{j*}_i может быть записана следующим образом

$$F^{j*}_i = \min_{\alpha,\ u} F(\alpha,\ u) \qquad \alpha \in G^j_i \qquad u \in U \qquad \text{(VI, 34)}$$

Поскольку $G^j_i \subset \overline{G}^j_i$, в соответствии с выражением (VI, 16) нижняя оценка ρ^j_i может быть получена решением следующей задачи:

$$\rho^j_i = \min_{\alpha,\ u} F(\alpha,\ u) \qquad \alpha \in \overline{G}^j_i \qquad u \in U \qquad \text{(VI, 35)}$$

В задаче (VI, 35) как u, так и α являются непрерывными переменными, и для ее решения могут быть использованы методы, изложенные в гл. III, IV.

Критерий прекращения ветвления. Пусть $\overline{\alpha}_{ij}$ — решение задачи (VI, 35). Легко видеть, что если $\overline{\alpha}_{ij}$ — целые, то выполняется условие (VI, 22) прекращения поиска. В случае, когда решение $\overline{\alpha}$ задачи (VI, 35) будет нецелым, $\overline{\overline{l}}$ схема, используемая в критерии (VI, 23), может, например, определяться следующим набором $\overline{\overline{\alpha}}$:

$$\overline{\overline{\alpha}}_{ij} = \begin{cases} 1 & \text{при} \quad \alpha^*_{ij} > 0{,}5 \\ 0 & \text{при} \quad \alpha^*_{ij} < 0{,}5 \end{cases} \qquad \text{(VI, 36)}$$

Остановимся теперь на вопросе — какой из структурных параметров α_{ij} целесообразно принимать в качестве параметра ветвления в вершине A_i^l. Наиболее очевидное правило — это выбор в качестве параметра ветвления такого параметра α_{ij}, который был бы наиболее близок к 0,5, т. е. выбор из условия

$$\min_{i,\, j} |0{,}5 - \alpha_{ij}|$$

Однако это правило имеет серьезный недостаток. Действительно, пусть, например, поток из k-го блока близок к нулю, тогда варьирование структурных параметров α_{jk}, $(j = \overline{1,\ N})$ близких к 0,5 будет очень мало влиять на критерий. Отсюда ветвление по этим параметрам может привести к схемам, оптимальные критерии для которых будут отличаться очень незначительно. Это может привести к большому числу лишних ветвлений. Поэтому, по-видимому, целесообразно придерживаться следующего правила: если структурный параметр соответствует потоку, близкому к нулю, то ветвление по нему производить не следует. В процессе поиска иногда можно существенно уменьшить число параметров α. Действительно, пусть в некоторой вершине A_i^l оказалось, что все параметры α_{ik}, $(k = \overline{1,\ N})$ равны нулю. Это значит, что на i-тый аппарат не подается ни один поток, и выход этого блока будет нулевым. Отсюда следует, что не имеет смысла варьировать параметры α_{ki}, $(k = \overline{1,\ N})$, связанные с выходным потоком этого блока, и все эти параметры можно приравнять нулю. После этого надо проверить, не появилось ли новых совокупностей α_{lj} $(j = \overline{1,\ N})$, равных нулю, и т. д.

Синтез ХТС при наличии фазовых ограничений

Будем исходить из предположения, что структурные параметры α_{ij} удовлетворяют условию (VI, 26), что в критерии (VI, 30) $P^{(k)} = 0$, $(k = \overline{1,\ N})$ и имеются фазовые ограничения (I,11). Структурные параметры 2-го рода δ введем следующим образом:

$$\delta_i = \begin{cases} 1, & \text{если } i\text{-ый блок включается в схему} \\ 0, & \text{если } i\text{-ый блок не включается в схему} \end{cases} \qquad \text{(VI, 37)}$$

Введем вспомогательный вектор γ, компоненты которого равны

$$\gamma_i = \begin{cases} 1, & \text{если в } i\text{-ом блоке учитываются фазовые ограничения} \\ 0, & \text{если они не учитываются} \end{cases} \qquad \text{(VI, 38)}$$

Перепишем соотношения (I, 1), (I, 11), (I, 15) в виде:

$$F = \sum_{k=1}^{N} \delta_k F^{(k)} (x^{(k)},\ u^{(k)},\ z^{(k)}) \qquad \text{(VI, 39)}$$

$$\overline{\varphi}^{(k)} (\delta,\ \gamma,\ x) \equiv \delta_k \gamma_k \varphi^{(k)} (x^{(k)}) \leqslant a^{(k)} \qquad \text{(VI, 40)}$$

$$z^{(k)} = \delta_k f^{(k)} (x^{(k)},\ u^{(k)}) \qquad \text{(VI, 41)}$$

Если положить $\delta_k = 0$, то из рассмотрения исключается k-тый блок, поскольку его вклад в общий критерий и выходные перемен-

ные будет равен нулю, а фазовое ограничение (VI, 40) будет заведомо выполняться. Если положить $\gamma_l = 0$, то в l-ом блоке фазовое ограничение заведомо выполняется. Если задать значения переменных α, δ, u, то рассчитав схему, можно найти соответствующее значение критерия F, отсюда F есть некоторая сложная функция этих переменных $F = F(\alpha, \delta, u)$. Задача синтеза с учетом фазовых ограничений может быть представлена следующей дискретно-непрерывной экстремальной задачей

$$\min_{\alpha, \delta, u} F(\alpha, \delta, u) \qquad u \in U \qquad \alpha \in \overline{G}_0^0 \qquad \overline{\varphi}^{(k)}(\delta, 1, x) \leqslant a \quad \text{(VI, 42)}$$

Задача синтеза схемы без учета фазовых ограничений выглядит следующим образом:

$$\min_{\alpha, u} F(\alpha, 1, u) \qquad \alpha \in \overline{G}_0^0 \qquad u \in U$$

Правило ветвления. Множество $A_{p_j}^j$ разбивается на два подмножества, в схемах одного из которых будет присутствовать некоторый (например, k-тый блок), а в схемах другого — отсутствовать (назовем этот блок блоком ветвления). Это значит, что схемы одного множества будут характеризоваться тем, что для них $\delta_k = 1$, для схем же другого множества $\delta_k = 0$. Введем вспомогательный вектор Γ_i^j, компоненты которого $\bar{\gamma}_i^j$ равны

$$\bar{\gamma}_k = \begin{cases} 1, & \text{если} \quad \delta_k \in \overline{D}_i^j \\ 0, & \text{если} \quad \delta_k \in \overline{\overline{D}}_i^j \end{cases} \qquad \text{(VI, 43)}$$

Задача определения нижней границы F_i^{j*} множества $\overline{A}_i^j$ может быть записана следующим образом:

$$F_i^{j*} = \min_{\alpha \in \overline{G}_0^0,\ \delta \in D_i^j,\ u \in U,\ \overline{\varphi}(\delta, 1, x) \leqslant a} F(\alpha, \delta, u) \qquad \text{(VI, 44)}$$

Покажем, что нижняя оценка ρ_i^j может быть получена решением следующей задачи:

$$\rho_i^j = \min_{\alpha \in \overline{G}_0^0,\ u \in U} F\left(\alpha, \Delta_i^j, u\right) \qquad \text{(VI, 45)}$$

$$\overline{\varphi}\left(\Delta_i^j, \Gamma_i^j, x\right) \leqslant a \qquad \text{(VI, 46)}$$

Заметим, что в то время, как в задаче (VI, 44) компоненты вектора δ являются поисковыми (двоичными) переменными, в задаче (VI, 45) — это фиксированные числа. Покажем, прежде всего, что множество схем, определяемое условиями $\delta = \Delta_i^j$, $\alpha \in G_0^0$, среди которых ищется оптимальная схема в задаче (VI,45) совпадает с множеством схем $\overline{A}_i^j$, среди которых ищется оптимальная схема в задаче (VI, 44). Действительно, возьмем произвольную схему, принадлежащую $\overline{A}_i^j$, которая характеризуется набором $\overline{\delta} \in D_i^j$, $\overline{\alpha} \in \overline{G}_0^0$. Ясно, что та же схема может быть получена и при выполнении условии $\delta = \Delta_i^j$,

$\alpha \in \overline{G}_0^0$. Для этого надо выбрать $\alpha_{ij} = \overline{\alpha}_{ij}$, $(j = \overline{1, N})$, если $\overline{\delta}_i \neq 0$ и $\alpha_{ij} = \alpha_{ji} = 0$, $(j = \overline{1, N})$, если $\overline{\delta}_i = 0$.

Рассмотрим теперь смысл неравенства (VI, 46). Согласно определению вектора Γ_i^j условие (VI, 46) означает, что в задаче (VI, 45) отсутствуют фазовые ограничения для блоков, которые не являются блоками ветвления в оптимальных вершинах пути $A_0^0 A_i^j$ и наоборот, учитываются все фазовые ограничения, соответствующие блокам, которые являются блоками ветвления в оптимальных вершинах пути $A_0^0 A_i^j$. Поскольку задача (VI, 45) отличается от задачи (VI, 44) только тем, что в ней отсутствуют некоторые фазовые ограничения, то в соответствии с уравнениями (VI, 16) при любых значениях двоичных параметров $\delta \in D_i^j$ существует соотношение (VI, 12) и величина ρ_i^j действительно является нижней оценкой. В задаче (VI,45) все поисковые переменные являются непрерывными. В качестве блока ветвления в множестве $\overline{A}_{p_j}^j$ надо будет выбирать один из блоков, для которого в процессе решения задачи (VI,45) ограничения (I, 11) оказались нарушенными.

Условие окончания процедуры поиска. Условие (VI, 22) будет выполнено, если при решении задачи (VI, 45) все ограничения (I, 11) будут выполнены. В качестве l-той схемы в выражении (VI, 23) может быть взята схема, которая будет получена в результате решения задачи:

$$\min_{\alpha,\, u} F\left(\alpha,\ \Delta_i^j,\ u\right) \qquad \alpha \in \overline{G}_0^0 \qquad \overline{\varphi}\left(\Delta_i^j,\ 1,\ x\right) \leqslant a$$

В отличие от задачи (VI, 45) здесь учитываются фазовые ограничения во всех блоках, которые не подвергались ветвлению в вершинах пути $A_0^0 A_i^j$.

Синтез схем при наличии постоянных составляющих критериев

Будем исходить из предположения, что фазовые ограничения (I,11) отсутствуют, а параметры α удовлетворяют условиям (VI, 26). Структурные параметры δ введем также, как и в предыдущем случае [см. выражение (VI, 37)]. Введем также вспомогательный вектор γ, компоненты которого определяются следующим образом:

$$\gamma_i = \begin{cases} 1, & \text{если в } i\text{-ом блоке учитывается постоянная составляющая критерия} \\ 0, & \text{если она не учитывается} \end{cases}$$

Соотношение (I, I) опять запишем в виде (VI, 41), а критерий (VI, 30) в виде

$$F = \sum_{k=1}^{N} \delta_k \left[F^{(k)}\left(x^{(k)},\ u^{(k)},\ z^{(k)}\right) + \gamma_k P^{(k)}\right]$$

Если положить $\delta_k = 0$, то из рассмотрения исключается k-тый блок; если же положить $\gamma_k = 0$, то в k-том блоке не учитывается постоянная составляющая критерия. Значение F есть некоторая

сложная функция переменных u, α, δ, γ : $F = F(\alpha, \delta, \gamma, u)$. Задача синтеза в данном случае может быть представлена следующей дискретно-непрерывной задачей:

$$\min_{\delta, u, \alpha} F(\alpha, \delta, 1, u) \qquad u \in U \qquad \alpha \in \overline{G}_0^0 \tag{VI, 47}$$

Правило ветвления будет таким же, как в предыдущем случае. Задача определения величины F_i^{j*} может быть записана следующим образом:

$$F_i^{j*} = \min_{\alpha, \delta, u} F(\alpha, \delta, 1, u) \qquad \alpha \in \overline{G}_0^0 \qquad \delta \in D_i^j \qquad u \in U \tag{VI, 48}$$

а величина ρ_i^j определяется решением задачи

$$\rho_i^j = \min_{\alpha \in \overline{G}_0^0,\, u \in U} F\left(\alpha, \Delta_i^j, \Gamma_i^j, u\right) \tag{VI, 49}$$

где Γ_i^j, Δ_i^j определяется соотношениями (VI, 43), (VI, 25). Действительно, пусть схема q, определяемая значениями α^*, δ^*, u^*, дает оптимальную величину критерия в условиях задачи (VI, 48): $F_i^{j*} = F(\alpha^*, \delta^*, 1, u^*)$. Поскольку множества схем, среди которых ищется оптимальная схема, в задачах (VI, 48) и (VI, 49) совпадают (доказывается этот факт так же, как и в предыдущем случае), среди схем, определяемых условиями задачи (VI, 49), также найдется схема q, которая определяется значениями $\alpha = \alpha^*$, $\delta = \Delta_i^j$. Значение критерия для этой схемы в обозначениях задачи (VI, 49) будет равно $F\left(\alpha^*, \Delta_i^j, \Gamma_i^j, u\right)$. Величины $F\left(\alpha^*, \delta^*, 1, u^*\right)$ и $F\left(\alpha^*, \Delta_i^j, \Gamma_i^j, u^*\right)$ определяют значения критерия для одной и той же схемы, в которой управляющие переменные принимают одни и те же значения, следовательно всегда будет выполняться условие

$$F\left(\alpha^*, \Delta_i^j, \Gamma_i^j, u^*\right) \leqslant F\left(\alpha^*, \delta^*, 1, u^*\right)$$

поскольку при вычислении обеих величин значения $F^{(k)}$ в выражении (VI, 30) будет совпадать, а при вычислении $F\left(\alpha^*, \Delta_i^j, \Gamma_i^j, u^*\right)$ для некоторых блоков постоянные составляющие $P^{(k)}$, $(P^{(k)} \geqslant 0)$ не будут учитываться. Таким образом, значение критерия задачи (VI, 49) для схемы q будет меньше, чем величина F_i^{j*}; отсюда, поскольку ρ_i^j заведомо не больше величины $F\left(\alpha^*, \Delta_i^j, 1, u^*\right)$ наше утверждение доказано.

Условие окончания поиска. В качестве схемы $\overline{\overline{l}}$, используемой в уравнении (VI, 23), может быть взята схема, которая будет получена в результате решения задачи

$$\min_{\alpha \in G_0^0,\, u \in U} F\left(\alpha, \Delta_i^j, 1, u\right) \tag{VI, 50}$$

Задача (VI, 50) отличается от задачи (VI, 49) тем, что в ней учитываются постоянные составляющие критерия (VI, 30) во всех блоках, из которых составлено множество схем $\overline{A}_i^j$.

Будем исходить из предположения, что критерий оптимизации имеет вид (VI, 30); $P^{(k)} = 0$, $(k = \overline{1, N})$, структурные параметры удовлетворяют (VI,26); фазовые ограничения (I, 11) отсутствуют. Будем называть первой группой выходные потоки блоков с номерами 1, . . ., m, а второй группой — выходные потоки блоков с номерами $m + 1$, . . ., N. Пусть потоки внутри одной группы могут смешиваться один с другим, потоки же из разных групп смешиваться не могут. Введем векторы δ и γ, с помощью которых перепишем соотношения связи (I, 6).

$$x^{(i)} = \delta_i \sum_{j=1}^{m} \alpha_{ij} z^{(j)} + (1 - \delta_i \gamma_i) \sum_{j=m+1}^{N} \alpha_{ij} z^{(j)} \qquad \text{(VI, 51)}$$

Выполняются соотношения

$$x^{(i)} = \begin{cases} \sum_{j=1}^{m} \alpha_{ij} z^{(j)}, \text{ если } \delta_i \gamma_i = 1 & \text{(VI, 52)} \\ \sum_{j=m+1}^{N} \alpha_{ij} z^{(j)}, \text{ если } \delta_i = 0,\ \gamma_i = 1 & \text{(VI, 53)} \\ \sum_{j=1}^{N} \alpha_{ij} z^{(j)}, \text{ если } \delta_i = 1,\ \gamma_i = 0 & \text{(VI, 54)} \end{cases}$$

В данном случае критерий (VI, 30) есть функция переменных α, δ, γ, u : $F = F(\alpha, \delta, \gamma, u)$. Поскольку на вход каждого блока должны подаваться потоки либо первой, либо второй группы, задача синтеза ХТС в данном случае может быть записана в виде (VI,47).

Правило ветвления. Множество $\overline{A}_{p_j}^{l}$ разбивается на два подмножества; в схемах одного из них на входе некоторого блока (пусть это будет k-тый блок) будут смешиваться потоки первой группы, а в схемах другого на входе этого блока будут смешиваться потоки только второй группы. Для всех схем первого подмножества $\delta_k = 1$, а для схем другого подмножества $\delta_k = 0$. Задачи определения величин F_l^{l*} и ρ_l^{l} могут быть записаны соответственно в виде (VI, 48), (VI, 49), где компоненты вектора Γ_l^{l} определяются соотношением (VI, 43). В соответствии со способом формирования векторов Δ_l^{l}, Γ_l^{l} для блоков, которые являлись блоками ветвления в «оптимальных» вершинах пути $A_0^0 A_l^{l}$, соотношения связи будут одинаковы в задачах (VI, 48), (VI, 49). Для блоков, которые не входят в эту группу, соотношения связи в задаче (VI, 48) будут иметь вид либо (VI,52), либо (VI, 53), в то же время, в задаче (VI, 49) эти связи будут иметь вид (VI, 54), т. е. в этой задаче они имеют более общий вид. Отсюда в соответствии с выражением (VI,16) всегда будет выполняться неравенство (VI, 12).

Условие окончания процедуры поиска. Легко видеть, что если при решении задачи (VI, 49) соотношения связи во всех блоках примут вид (VI, 52), либо (VI, 53), то будет выполняться условие

(VI, 22). В случае, когда условия (VI, 52), (VI, 53) выполняются не во всех блоках, схема $\bar{l}$, используемая в выражении (VI, 23), будет определяться следующим набором δ:

$$\delta_i = \begin{cases} 1, \text{ если } \sum_{j=1}^{m} \alpha_j^* > \sum_{j=m+1}^{N} \alpha_{ij}^* \\ 0, \text{ если } \sum_{j=1}^{m} \alpha_{ij} < \sum_{j=m+1}^{N} \alpha_{ij}^* \end{cases}$$

где α^* — решение задачи (VI,49).

Синтез теплообменных систем

Пусть имеется множество N горячих потоков $S_h = \{S_{h1}, \ldots, S_{hN}\}$ и множество M холодных потоков $S_c = \{S_{c1}, \ldots, S_{cM}\}$. Для определенности будем считать, что $M \leqslant N$. Теплообменная система (ТС) состоит из совокупности рекуперативных теплообменников (в дальнейшем будем называть их теплообменниками) и совокупности нагревателей (греющий пар) и холодильников (охлаждающая вода); первая совокупность есть внутренняя система, вторая — внешняя система. Пусть начальные температуры горячих S_{hi}, $(i = \overline{1,\ N})$ и холодных S_{cj}, $(j = \overline{1,\ M})$ потоков равны, соответственно

$$T_{hi}^0\ (i = \overline{1,\ N}) \qquad T_{cj}^0\ (j = \overline{1,\ M}) \tag{VI, 55}$$

ТС должна понизить температуры горячих потоков до T_{hi}^k и повысить температуры холодных потоков до T_{cj}^k:

$$T_{hi}^k\ (i = \overline{1, N}) \qquad T_{cj}^k\ (j = \overline{1,\ M}) \tag{VI, 56}$$

Внешняя снстема строится следующим образом. Нагреватели (холоднльннкн) ставятся на холодных (горячнх) потоках, если за счет рекуперацнн тепла во внутренней системе не удается достнчь желаемых температур (VI, 56). Часто ТС стронтся таким образом, что на каждом потоке блоки внешней снстемы стоят после блоков внутренней системы (рнс. 39).

Температуру охлаждающей воды до н после i-го холоднльника обозначнм соответственно через T_u, $T_{ui}^{(k)}$. Допустнм, что начальные температуры охлаждающей воды у всех холоднльников одннаковы.

Будем нсходнть нз предположения, что нспользуются трубчатые противоточные теплообменннкн. Математическая модель теплообменника, в котором обмениваются теплом горячий S_{hi} и холодный S_{cj} потокн, имеет вид [135]

$$Q_{i,\,j} = W_i c_{pi} (T_{hi}^0 - T'_{hi}) = W_j c_{pj} (T'_{cj} - T_{cj}^0) \tag{VI, 57}$$

$$Q_{i,\,j} = \gamma A_{i,\,j} \frac{(T_{hi}^0 - T'_{cj}) - (T'_{hi} - T_{cj}^0)}{\ln [(T_{hi}^0 - T'_{cj})/(T'_{hi} - T_{cj}^0)]} \tag{VI, 58}$$

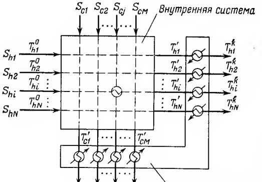


Рис. 39. Схема теплообменной системы.

где T'_{hi}, T'_{cj} — температуры горячего и холодного потоков на выходе теплообменника; γ — коэффициент теплопередачи; W_i, W_j — расходы i-го горячего и j-го холодного потоков; c_{pi} — теплоемкость i-го потока; $Q_{i,j}$ — количество тепла, передаваемого в теплообменнике, осуществляющим теплообмен между i-ым горячим и j-ым холодным потоками; $A_{i,j}$ — поверхность теплообмена.

Математическая модель холодильника также имеет вид (VI, 57), (VI, 58). Математическая модель нагревателя зависит от его типа. Если исключить величину Q_{ij} из системы (VI, 57), (VI, 58), получим систему из двух нелинейных уравнений, связывающую пять переменных

$$T^0_{hi} \qquad T'_{hi} \qquad T^0_{cj} \qquad T'_{cj} \qquad A_{i,j} \tag{VI, 59}$$

Задав любые три из них, можно найти остальные, решая систему из двух уравнений с двумя неизвестными.

Остановимся теперь на поисковых переменных ТС: теплообменнику будет соответствовать одна поисковая переменная (поверхность теплообмена $A_{i,j}$), холодильнику две (поверхность теплообмена A_{hi} и расход воды U_i, i — номер горячего потока, на котором стоит холодильник), нагревателю — одна (поверхность теплообмена A_{cj}, j — номер холодного потока, на котором стоит нагреватель). В нагревателе расход пара обычно выбирается из условия, равенства количества передаваемого тепла теплоте конденсации пара. Максимальное число поисковых переменных R в ТС будет равно

$$R = p + M + 2N$$

В критерии оптимизации F будем учитывать капитальные затраты и затраты на охлаждение и нагревание. Обозначим через M_p совокупность p пар чисел (i, j) (по числу теплообменников в ТС), в которой каждая пара (q, r) соответствует теплообменнику, осуществляющему теплообмен между q-тым горячим и r-тым холодными потоками. Тогда критерий оптимизации может быть записан в виде:

$$F = \sum_{(i,\,j) \in M_p} F^{(1)}_{i,\,j} + \sum_{k=1}^{N} F^{(2)}_k + \sum_{l=1}^{M} F^{(3)}_l \tag{VI, 60}$$

где $F^{(1)}_{(i,\,j)}$, $F^{(2)}_k$, $F^{(3)}_l$ затраты на теплообменник, холодильник и нагреватель, соответственно

$$F^{(1)}_{i,\,j} = \delta a A^b_{i,\,j};\; F^{(2)}_k = \delta a A^b_{hk} + \alpha_1 U_k;\; F^{(3)}_l = \delta a A^b_{cl} + \beta_1 V_l \qquad \text{(VI, 61)}$$

δ — коэффициент окупаемости капитальных вложений; α_1, β_1 — цены единицы массы охлаждающей воды и пара, соответственно; a, b — некоторые корреляционные коэффициенты (обычно $0{,}6 \leqslant b \leqslant 1$); будем исходить из предположения, что цены на воду (пар) не зависят от номера холодильника (нагревателя).

Задача синтеза ТС ставится следующим образом [117], [126], [135]—[137]; найти структуру внутренней системы, в случае необходимости поставить холодильники на горячих потоках и нагреватели на холодных, определить такие значения всех поисковых технологических переменных, при которых температуры горячих и холодных потоков до и после ТС принимают значения (VI, 55), (VI, 56), а критерий (VI, 60) — минимальное значение. Из постановки задачи виден ее компромиссный характер по отношению к капитальным и эксплуатационным затратам. Так, добиться понижения температур горячих потоков и повышения температур холодных потоков можно, поставив на каждом горячем потоке холодильник, а на каждом холодном — нагреватель. Этому варианту схемы соответствуют максимальные эксплуатационные расходы. Другой вариант схемы соответствует максимальной рекуперации тепла, при которой эксплуатационные расходы будут минимальными. Оптимальная же структура ТС будет соответствовать некоторому компромиссному варианту.

Синтез систем теплообмена на основе задачи о назначении

Одной из первых работ по синтезу ТС на основе задачи о назначениях была работа [138]. В дальнейшем этот метод был развит в работах [139], [140]. В дальнейшем произвольную ТС, построенную для совокупностей горячих S_h и холодных S_c потоков будем обозначать через ТС ($S_h \times S_c$). ТС будем называть б а з о в о й, если она имеет структуру, показанную на рис. 39, и каждый поток во внутренней системе может обмениваться теплом только один раз. Задачу синтеза базовой ТС ($S_h \times S_c$) будем называть о с н о в н о й з а д а ч е й с и н т е з а размерности $N \times M$.

Синтез базовой теплообменной системы. Рассмотрим вначале случай, когда числа холодных и горячих потоков равны $N = M$, а число теплообменников также равно N. Покажем, что при сделанных предположениях задача синтеза ТС может быть сведена к специальной задаче целочисленного линейного программирования — задаче о назначениях [127, с. 405]. Введем двоичные переменные x_{ij} следующим образом:

$$x_{ij} = \begin{cases} 1, & \text{если есть теплообмен между горячим потоком } S_{hi} \\ & \text{и холодным потоком } S_{cj} \\ 0, & \text{если нет теплообмена между горячим потоком } S_{hi} \\ & \text{и холодным потоком } S_{cj} \end{cases} \qquad \text{(VI, 62)}$$

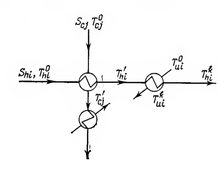


Рис. 40. Схема ЭБС $(i,\ j)$

Матрица $X = \|x_{ij}\|$ называется **матрицей назначения**. Поскольку в базовой ТС на каждом горячем и холодном потоке может стоять только один теплообменник, переменные x_{ij} должны удовлетворять следующим соотношениям:

$$\sum_{j=1}^{N} x_{ij} = 1 \qquad \sum_{i=1}^{N} x_{ij} = 1 \quad \text{(IV, 63)}$$

Из выражений (VI, 62), (VI, 63) вытекает, что в каждой строке (столбце) матрицы X имеется только один элемент равный единице, а все остальные элементы строки (столбца) равны нулю. Поскольку стоимость и начальная температура охлаждающей воды (пара) не зависит от номера холодильника (нагревателя), отпадает проблема перебора при построении внешней системы, и за данным горячим (холодным) потоком может быть закреплен любой холодильник (нагреватель). Поэтому закрепим за теплообменником, в котором обмениваются теплом потоки S_{hi}, S_{cj}, один холодильник, который будет обеспечивать охлаждение горячего потока S_{hi} до температуры T^k_{hi} и один нагреватель, который будет обеспечивать нагревание холодного потока S_{cj} до температуры T^k_{cj} (рис. 40). В данном случае полученная совокупность теплообменника, нагревателя и холодильника будет элементарным блоком синтеза, который обозначим через ЭБС $(i,\ j)$. Любая базовая ТС состоит из N таких ЭБС $(i,\ j)$. В связи с этим перепишем критерий (VI, 60), учитывая, что $p = M = N$

$$F = \sum_{(i,\ j) \in M_N} F_{i,\ j}$$

где $F_{i,\ j} = F^{(1)}_{i,\ j} + F^{(2)}_i + F^{(3)}_j$ — часть критерия (VI,60), относящаяся к ЭБС $(i,\ j)$. Задача оптимизации ЭБС $(i,\ j)$ с критерием $F_{i,\ j}$ записывается следующим образом

$$\min_{A_{i,\ j},\ A_{cj},\ A_{hi},\ U_i} F_{i,\ j} \quad \text{(VI, 64)}$$

при условии

$$T'_{cj} = T^k_{cj} \qquad T'_{hi} = T^k_{hi} \quad \text{(VI, 65)}$$

Поскольку в базовой ТС на каждом потоке стоит только один теплообменник, между элементарными блоками синтеза нет взаимного влияния. Так что каждый ЭБС можно оптимизировать отдельно; при этом будет получен оптимальный режим для всей ТС, т. е. будет справедливо равенство

$$\min_{A_{i,\ j},\ A_{cj},\ A_{hi},\ U_i} F = \sum \min_{A_{i,\ j},\ A_{cj},\ A_{hi},\ U_i} F_{i,\ j}$$

при условии выполнения соотношений (VI, 65) для $(i, j = \overline{1, N})$. Обозначим через f_{ij} оптимальное значение критерия $F_{i,j}$, полученное решением задачи (VI, 64), (VI, 65).

$(N \times N)$-Матрицу $\Phi = \|f_{ij}\|$ будем называть матрицей оценок. Введем целевую функцию следующим образом:

$$F^{(1)} = \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} x_{ij} \qquad \text{(VI, 66)}$$

В соответствии с условиями (VI,62), (VI, 63) в $F^{(1)}$ в любом случае будет присутствовать только N каких-либо величин f_{ij} $(i, j = \overline{1, N})$. Задача оптимизации базовой ТС может быть теперь записана следующим образом:

$$\min_{x_{ij}} \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} x_{ij} \qquad \text{(VI, 67)}$$

$$\sum_{j=1}^{N} x_{ij} = 1 \qquad \sum_{i=1}^{N} x_{ij} = 1 \qquad \text{(VI, 68)}$$

Это специальная задача целочисленного линейного программирования, которая носит название задачи о назначениях. Для нее имеются хорошо разработанные методы решения. Решив эту задачу; найдем оптимальную структуру базовой ТС.

Рассмотрим теперь случай, когда числа холодных и горячих потоков не равны $(N > M)$, а число теплообменников равно M. При этом по-прежнему на каждом холодном потоке обязательно будет стоять теплообменник. В то же время, на M горячих потоках будет стоять теплообменник, а на остальных $N - M$ горячих потоках он будет отсутствовать. Двоичные переменные x_{ij} в данном случае удовлетворяют соотношениям

$$\sum_{j=1}^{M} x_{ij} \leqslant 1, \ (i = \overline{1, N}) \qquad \sum_{i=1}^{N} x_{ij} = 1, \ (j = \overline{1, M}) \qquad \text{(VI, 68 а)}$$

Если на p-том горячем потоке теплообменник отсутствует, то охлаждение этого потока будет осуществляться только за счет холодильника. Обозначим через f_i минимальные затраты на холодильник, охлаждающий i-тый горячий поток, в том случае, когда отсутствует теплообмен этого потока с холодными потоками. Введем величины

$$\varphi_i = \sum_{j=1}^{M} f_{ij} x_{ij} + \left(1 - \sum_{j=1}^{M} x_{ij}\right) f_i$$

При этом выполняется соотношение

$$\varphi_i = \begin{cases} f_{ij}, \text{ если } i\text{-тый горячий поток обменивается теплом с } j\text{-тым холодным} \\ f_i, \text{ если } i\text{-тый горячий поток не обменивается теплом} \\ \quad \text{ни с одним холодным потоком} \end{cases}$$

Действительно, если i-тый горячий поток обменивается теплом с j-тым холодным потоком, то с учетом ограничений (VI,68а) $x_{ij} =$

$= 1$; $x_{ip} = 0$ $(p \neq j)$. Отсюда второе слагаемое в выражении для φ_i равно нулю и $\varphi_i = f_{ij}$. Если i-тый горячий поток не обменивается теплом с холодными потоками, то $x_{ij} = 0$, $(j = \overline{1, M})$ и первое слагаемое в φ_i равно нулю, а второе равно f_i. Таким образом, величина f_i представляет собой стоимость охлаждения i-го горячего потока при любом способе охлаждения, а величина суммы всех φ_i $(i = 1, \ldots, n)$ при наличии ограничений (VI,68 а)—стоимость одной из возможных ТС. Так что задача синтеза ТС запишется следующим образом:

$$\min_{x_{ij}} \sum_{i=1}^{N} \left[\sum_{j=1}^{M} f_{ij} + \left(1 - \sum_{j=1}^{M} x_{ij}\right) f_i \right]$$

при условии выполнения соотношения (VI, 68). После приведения подобных членов в выражении для критерия получим несимметричную задачу о назначении.

Синтез ТС произвольной структуры. Рассмотрим два подхода к этой задаче. При первом подходе так же, как и в работах [138]—[140] каждый горячий S_{hi} (холодный S_{cj}) поток разбиваем на n_i (m_j) элементарных потоков таким образом, чтобы выполнялись следующие условия:

а) массовый расход любого элементарного потока равен массовому расходу исходного горячего (холодного) потока;

б) начальная температура $T^0_{hi,\,l+1}$ $(l + 1)$-го элементарного потока равна конечной температуре $T^k_{hi,\,l}$ l-го элементарного потока

$$T^k_{cj,\,r} - T^0_{cj,\,r+1} = 0 \qquad T^k_{hi,\,l} - T^0_{hi,\,l+1} = 0$$

$$l = \overline{1,\ n_i - 1};\ r = \overline{1,\ m_j - 1} \qquad \text{(VI, 69)}$$

в) начальная температура первого элементарного потока равна начальной температуре исходного потока, а конечная температура последнего элементарного потока равна конечной температуре исходного потока

$$T^0_{hi,\,1} = T^0_{hi} \qquad T^k_{hi,\,n_i} = T^k_{hi} \qquad T^0_{cj,\,1} = T^0_{cj} \qquad T^k_{cj,\,m_j} = T^k_{ci} \qquad \text{(VI, 70)}$$

Так же, как в работах [138]—[140] будем исходить из предположения, что каждый элементарный горячий (холодный) поток может обмениваться теплом только один раз. Однако, в отличие от этих работ мы не будем предполагать, что от каждого элементарного потока отбирается, а каждому холодному элементарному потоку добавляется одно и то же количество тепла q. (Способы выбора величины q даны в работах [138]—[140]). Прежде всего это предположение является довольно ограничительным, поскольку ниоткуда не следует, что в каждом теплообменнике оптимальной ТС количество обмениваемого тепла должно быть одним и тем же. Кроме того, это предположение запрещает теплообмен между потоками, между которыми он физически допустим, но будет меньше, чем q [141].

Дадим всем горячим элементарным потокам новую (сплошную) нумерацию и будем рассматривать их как новую совокупность горячих потоков $\overline{S}_h$. Аналогичным образом, все холодные элементарные потоки будем рассматривать как новую совокупность холодных потоков $\overline{S}_c$. Числа $\overline{N}$, $\overline{M}$ потоков в множествах $\overline{S}_h$, $\overline{S}_c$ будут равны $\overline{N} = n_1 + \cdots n_N$; $\overline{M} = m_1 + \cdots m_M$. Воспользуемся декомпозиционным принципом закрепления и зафиксируем начальные и конечные температуры всех элементарных потоков при соблюдении условий (VI, 69), (VI, 70). Для новых множеств $\overline{S}_h$, $\overline{S}_c$ горячих и холодных потоков построит оптимальную базовую ТС ($\overline{S}_h \times \overline{S}_c$); для этого мы должны решить основную задачу синтеза размерности $\overline{N} \times \overline{M}$. Конечно, решение этой задачи не дает точного решения первоначальной задачи синтеза ТС, поскольку начальные и конечные температуры элементарных потоков были заданы произвольно. В связи с этим предлагается следующая двухуровневая процедура. На первом уровне решается основная задача синтеза ТС размерности $\overline{N} \times \overline{M}$ при заданных начальных и конечных значениях температур элементарных потоков. На втором уровне проводится оптимизация схемы найденной (фиксированной) структуры, при этом в качестве поисковых используются все технологические переменные $A_{i,j}$, A_{hi}, A_{cj}, U_i. В результате решения этой задачи будут уточнены значения начальных и конечных температур всех элементарных потоков. После этого опять решается основная задача синтеза ТС при новых значениях начальных и конечных температур элементарных потоков, и т. д.

Подытожим — какие основные операции приходится проводить на каждой итерации при использовании этого метода. Для простоты будем считать, что $N = M$ и что каждый исходный поток разбит на n элементарных потоков, т. е. $\overline{N} = nN$. На нижнем уровне необходимо будет произвести следующие операции:

1) для формирования матрицы Ф решить $\overline{N}^2 = n^2N^2$ задач оптимизации ЭБС (i, j), каждая из которых сводится к оптимизации критерия $F_{i,j}$ по четырем переменным $A_{i,j}$, A_{hi}, A_{cj}, U_i;

2) решить задачу о назначениях с ($nN \times nN$)-матрицей.

На верхнем уровне необходимо будет решить задачу оптимизации ТС с фиксированной структурой и числом теплообменников, равным nN. При этом число поисковых переменных будет равно $4nN$ ($3nN$ — поверхности теплообменников, нагревателей и холодильников, nN — расходы воды).

Решение задачи определения элемента f_{ij} матрицы оценок Ф может быть существенно упрощено, если поверхности теплообмена в теплообменнике и холодильнике выбирать из условия максимального сближения температур. Для теплообменника это условие имеет вид

$$T^0_{hi} - T'_{cj} = \Delta T_{\min} \tag{VI, 71}$$

где $\Delta T_{\min}$ заданная величина. Аналогичное условие может быть выписано для холодильника. Таким образом, к двум условиям

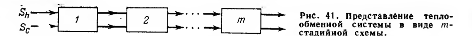


Рис. 41. Представление теплообменной системы в виде m-стадийной схемы.

(VI, 65) в ЭБС (i, j) в данном случае добавляются еще два условия, и для определения f_{ij} необходимо будет решить систему из четырех уравнений с четырьмя неизвестными $A_{i,j}$, A_{cj}, A_{hi}, U_i. Легко показать, что решение указанной системы сведется к последовательному расчету теплообменника, холодильника и нагревателя. Ясно, что при этом мы не получим оптимального значения величины F.

Рассмотрим теперь второй подход. Представим вначале ТС в виде m-стадийной схемы (рис. 41), каждая стадия которой представляет собой базовую ТС $(S_h \times S_c)$. Горячие и холодные потоки последовательно проходят все стадии с 1-й по m-ю. Ясно, что полученная ТС будет глобальной, из которой может быть получена любая ТС. Положив нулю значения поверхности теплообмена у всех теплообменников кроме одного на каждой стадии, придем к многостадийной схеме, рассмотренной в работе [142].

Рассмотрим подход к синтезу ТС, использующий построенную глобальную ТС. Он также основывается на декомпозиционном принципе закрепления, сводящим задачу синтеза ТС к двухуровневой оптимизационной процедуре. В соответствии с принципом закрепления закрепим в m-стадийной схеме температуры всех горячих и холодных промежуточных потоков. Рассмотрим k-ую стадию $(k \leqslant m)$. На этой стадии имеется совокупность S_h горячих и S_c холодных потоков, с известными входными и выходными температурами. Определим наилучшую ТС для k-той стадии. Поскольку k-тая стадия представляет собой базовую ТС $(S_h \times S_c)$, задача синтеза ТС k-той стадии сводится к основной задаче синтеза размерности $N \times M$. Решив эту задачу для всех стадий глобальной схемы, найдем некоторую структуру ТС, что будет являться окончанием процедуры 1-го уровня. На втором уровне температуры всех промежуточных потоков освобождаются от закрепления и проводится оптимизация всей ТС, при этом поисковыми переменными являются все технологические параметры. Поскольку все переменные здесь непрерывные, на этом уровне используется один из поисковых методов. После окончания оптимизации будут получены новые значения температур для промежуточных потоков. Закрепим их на этих значениях и опять перейдем к решению задач 1-го уровня. Преимущество этого подхода к построению ТС перед предыдущим состоит в том, что решение одной задачи о назначениях большой размерности на 1-м уровне заменяется решением m задач о назначениях меньшей размерности. Однако, этот подход нельзя применять для синтеза циклических ТС.

Некоторые обобщения. I. Первый и второй подходы легко распространить на случай, когда необходимо синтезировать теплообменную систему с разветвлением потоков. Для примера рассмотрим первый подход. Пусть описаным способом введены элементарные потоки. Назовем их элементарными потоками 1-го рода. Рассмотрим l-тый элементарный поток, полученный из горячего потока S_{hi}.

Рис. 42. Схема разбиения элементарного потока 1-го рода на совокупность элементарных потоков 2-го рода:

а — исходный технологический поток; *б* — совокупность элементарных потоков 2-го рода, полученных из *l*-го элементарного потока 1-го рода.

Его расход равен W_i.·С помощью делителя потоков (рис. 42) разобъем этот элементарный поток 1-го рода на m_{il} элементарных потоков 2-го рода. Полученную совокупность обозначим через S_{hi}^{l}. Каждый q-тый элементарный поток 2-го рода будет характеризоваться расходом $\alpha_q^{il} W_i$ при

$$\sum_{q=1}^{m_{il}} \alpha_q^{il} = 1$$

Начальная же и конечная температуры будут совпадать с начальной и конечной температурами элементарного потока 1-го рода, т. е. будут равны $T_{hi,\,l}^{0}$ и $T_{hi,\,l}^{k}$, соответственно.

Проделаем такую же операцию со всеми элементарными потоками, при этом опять будет исходить из того, что горячий элементарный поток 2-го рода может обмениваться теплом только один раз с любым холодным элементарным потоком 2-го рода, и наоборот. После теплообмена все элементарные потоки 2-го рода, принадлежащие одной и той же совокупности S_{hi}^{l} (или S_c^{jm}) объединяются в смесителе. Объединим все горячие элементарные потоки 2-го рода в новую совокупность горячих потоков $\overline{\overline{S}}_h$, а холодные — в новую совокупность $\overline{\overline{S}}_c$. После этого опять можно воспользоваться двухуровневой процедурой первого подхода. Отличие будет состоять только в том, что в качестве поисковых переменных помимо начальных и конечных температур элементарных потоков будут использоваться также структурные параметры α_q^{il}. При этом размерность задач оптимизации как первого, так и второго уровня может существенно возрасти.

II. Обычно стоимость транспортировки потоков между теплообменниками мала по сравнению с затратами на охлаждение и нагревание. Однако для вязких или коррозионноактивных веществ стоимость транспортировки может стать сравнимой с остальными компонентами критерия (VI, 61) [143]. В этом случае при постановке задачи синтеза ТС необходимо учитывать не только структуру, но и расположение теплообменников. В этом случае при определении элементов f_{if} матрицы Ф необходимо учитывать также стоимость труб, связывающих теплообменник с источниками холодного и горячего потоков, а также стоимость перекачивания потоков.

III. Задача синтеза ТС решалась в идеализированной постановке. Так, коэффициент теплопередачи γ может быть рассчитан, если известны некоторые параметры, которые не участвовали в описанной процедуре (например, диаметр трубок и др.). Поэтому после того, как описанная процедура синтеза ТС будет проведена, необ-

ходимо уточнить коэффициенты теплопередачи, а затем заново провести описанную выше процедуру синтеза и т. д.

Некоторые параметры такие как диаметр, длина и число трубок, могут принимать только дискретные значения. Поэтому, после того, как найдено значение поверхности теплообмена для данного теплообменника, необходимо подобрать эти величины так, чтобы максимально приблизиться к этому значению. Другими словами, в этом случае дискретная экстремальная задача сначала заменяется непрерывной, а решение непрерывной задачи приближается дискретным решением. Можно однако пойти другим путем: формируя элементы матрицы Ф, в качестве поисковых переменных использовать не поверхности теплообмена, а диаметр, длину и число трубок.

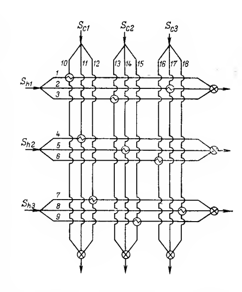
Метод структурных параметров

Для применения описанной выше стандартной процедуры метода структурных параметров на все выходные потоки теплообменников следует поставить делители потоков, а на входные потоки — смесители потоков. Кроме того, делители потоков должны быть поставлены на все входные холодные и горячие потоки. Естественно, что в смесителе, стоящем на входном горячем потоке теплообменника, должны смешиваться только горячие потоки, а в смесителе, стоящем на входном холодном потоке,— только холодные потоки. Пусть число теплообменников равно m тогда число структурных параметров для случая, когда рассматриваются только системы без обратных связей, будет равно $m\,(m-1)+(M+N)\,m$. Однако стандартная процедура метода структурных параметров имеет существенный недостаток. Дело в том, что фактически не разрешается смешивать потоки. Это условие будет выполнено, если наложить требование, чтобы все структурные параметры были целыми числами. Это требование, как мы видели, приводит к необходимости проведения трудоемкой процедуры метода «ветвей и границ» совместно с методом структурных параметров.

Опишем другой подход к построению глобальной схемы, в которой заранее гарантируется несмешиваемость исходных потоков. Будем исходить из предположения, что ТС имеет структуру, показанную на рис. 39. Представим внутреннюю систему в виде m-стадийной схемы (см. рис. 41). При этом k-тую стадию ($m \leqslant k$) построим следующим образом. Поставим на каждом горячем и холодном входном потоке k-той стадии делитель потока. Тот из них, который стоит на горячем потоке S_{hi} будет делить поток на M частей (по числу холодных потоков); эти потоки будет называть элементарными. Каждый делитель, стоящий на холодном потоке S_{cj} будет делить его на N элементарных потоков (по числу горячих потоков). Общее число горячих (холодных) потоков будет равно NM. Построим ТС, которая будет обладать следующими свойствами:

1) каждый элементарный поток может обмениваться теплом только один раз; 2) при рассмотрении двух совокупностей элементарных потоков, соответствующих исходному горячему потоку S_{hi}

Рис. 43. Отдельная стадия глобальной схемы теплообменной системы для трех горячих и трех холодных потоков:
1—9 — номера горячих элементарных потоков; 10—18 — то же, холодных.



и холодному потоку S_{cj}, теплообмен может проводиться только между одной парой горячих и холодных элементарных потоков из этих совокупностей.

На данной стадии будет NM теплообменников и после прохождения системы теплообменников все элементарные потоки, соответствующие исходному горячему (холодному) потоку, подаются в смеситель. Такой принцип построения глобальной схемы обеспечивает с одной стороны возможность теплообмена на данной стадии между любыми исходными горячим и холодным потоками, а с другой — их несмешиваемость. На рис. 43 приведена одна из стадий глобальной схемы для случая, когда имеется 3 горячих и 3 холодных исходных потока. Последняя $(m + 1)$-я стадия соответствует внешней системе (см. рис. 39). Легко видеть, что данной стадии соответствует $2NM$ структурных параметров. А всего в $(m + 1)$-стадийной глобальной схеме будет $2mNM$ структурных параметров. Число технологических поисковых переменных будет равно $mNM + 2N + M$; таким образом общее число R поисковых переменных будет равно

$$R = 3mNM + 2N + M \tag{VI, 72}$$

Как обычно, структурные параметры являются непрерывными переменными, удовлетворяющими соотношениям (I, 7), (VI, 26). Давая структурным параметрам определенные значения, можно из глобальной получить любую заданную ТС (без рециклов), а после проведения оптимизации глобальной схемы, получить схему ТС, наилучшую из всех возможных. Поскольку в глобальной схеме все поисковые переменные (структурные и технологические) непрерывны, для ее оптимизации могут быть использованы численные методы нелинейного программирования. После решения задачи оптимизации глобальной схемы ТС будут получены какие-то значения структурных параметров (вообще говоря, нецелые). Однако, если условия задачи разрешают разветвления потоков, это не страшно: если структурные параметры, соответствующие какому-либо потоку, окажутся нецелыми, на нем надо ставить делитель потоков. Если же разветвление потоков не разрешается, необходимо потребовать целочисленность структурных параметров. В этом случае, также как и при использовании обычной глобальной схемы,

придется применять процедуру метода «ветвей и границ» совместно с методом структурных параметров. К сожалению, этот способ построения глобальной схемы, так же, как и обычный, имеет серьезный недостаток, обусловленный с многоэкстремальностью критерия [144] и большой размерностью задачи. Все это делает затруднительным использованного данного подхода.

Сравним теперь 1-й и 2-й подходы с методом структурных параметров. Будем считать, что $N = M$ и что число стадий m в глобальной схеме ТС, используемой в методе структурных параметров и во 2-м подходе, равно числу n элементарных потоков, на которые разбивают исходные потоки в 1-м подходе. Тогда при использовании метода структурных параметров задача синтеза ТС сведется к задаче нелинейного программирования с числом переменных $R = 3nN^2 + 3N$. При использовании 1-го подхода на каждой итерации потребуется решить n^2N^2 задач оптимизации размерности 4 и одну задачу оптимизации размерности $4nN$. При использовании 2-го подхода на каждой итерации придется решить nN^2 задач оптимизации размерности 4 и одну задачу размерности $4nN$. Так, если $n = 3$, $N = M = 5$, то размерность R задачи нелинейного программирования при использовании метода структурных параметров будет равна 390. При использовании 1-го подхода на каждой итерации придется решить 275 экстремальных задач размерности 4 и одну задачу размерности 60. При использовании 2-го подхода на каждой итерации придется решить 75 экстремальных задач размерности 4 и одну задачу размерности 60.

Однако преимущество 1-го и 2-го подходов состоит не только в уменьшении размерности экстремальных задач, но и связано с проблемой многоэкстремальности. Метод структурных параметров приводит обычно к многоэкстремальной задаче [122], что связано, по-видимому, с тем, что в глобальную схему включены все возможные варианты схем ТС. Выбор той или иной структуры определяется решением задачи нелинейного программирования. В то же время при 1-м и 2-м подходах основная тяжесть выбора структуры ложится на решение задачи о назначениях, а с помощью метода нелинейного программирования приходится решать задачу оптимизации ТС, фиксированной структуры. Конечно, полностью избавиться от многоэкстремальности не удается, поскольку даже задача оптимизации ТС фиксированной структуры часто оказывается многоэкстремальной.

Синтез теплообменной системы как части химико-технологической схемы произвольной структуры

При выводе рассмотренных выше методов синтеза ТС существенно использовались следующие предположения.

1. Расходы на перекачку потоков через ТС не учитывались.
2. Значения начальных и конечных температур холодных и горячих потоков (VI, 55), (VI, 56) считаются заданными.

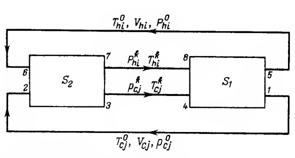


Рис. 44. Схема неоднородной ХТС:
S_1 — гетерогенная система; S_2 — теплообменная система; *1*—*8* — номера потоков.

Избавиться от этих упрощающих предположений можно, рассмотрев совместную работу теплообменной системы с остальной частью схемы. В связи с этим рассмотрим некоторую схему S (рис. 44), состоящую из теплообменной системы S и произвольной части S_1, в которой имеется какое-то число компрессоров. Подсистемы S_1 и S_2 связаны холодными (*1*—*2*), (*3*—*4*) и горячими (*5*—*6*), (*7*—*8*) потоками. Обозначим температуры горячих и холодных потоков на входе и выходе ТС так, как это сделано в формулах (VI, 55), (VI, 56), а соответствующие давления через

$$p^0_{hi},\ p^0_{cj}\ \text{и}\ p^k_{hi},\ p^k_{cj}\quad (i = \overline{1,\ N};\ j = \overline{1,\ M}) \qquad \text{(VI, 73)}$$

Расходы горячих и холодных потоков обозначим через

$$V_{hi},\ V_{cj}\qquad (i = \overline{1,\ N};\ j = \overline{1,\ M}) \qquad \text{(VI, 74)}$$

Пусть задача состоит в определении оптимального режима схемы S с одновременным синтезом ТС S_2. Рассмотрим вначале вспомогательную задачу синтеза ТС, в которой будут заданы не только температуры (VI, 55), (VI, 56), но и давления (VI, 73) на входе и выходе. Для простоты будет считать, что в ТС является базовой. В данном случае в качестве поисковых переменных при решении задачи оптимизации ЭБС (i, j) необходимо использовать длину l, диаметр d и число трубок n. Буквы т, н, х в нижнем индексе этих переменных будут означать, что данная величина относится к теплообменнику, нагревателю или холодильнику соответственно. Обозначим через $\Delta p_{\text{г. г}\,i}$, $\Delta p_{\text{т. х}\,j}$, $\Delta p_{\text{х}}$, $\Delta p_{\text{н}}$ — перепады давлений по i-му горячему и j-му холодному потоку в теплообменнике, холодильнике и нагревателе соответственно. Ясно, что эти величины являются функциями соответствующих величин l, d и n. При решении задачи оптимизации ЭБС (i, j) в данном случае появятся следующие дополнительные ограничения:

$$\begin{aligned} &\Delta p_{\text{т. г.}\,i}\,(l_{\text{т}},\ d_{\text{т}},\ n_{\text{т}}) + \Delta p_{\text{х}}\,(l_{\text{х}},\ d_{\text{х}},\ n_{\text{х}}) = p^0_{hi} - p^k_{hi} \\ &\Delta p_{\text{т. х.}\,j}\,(l_{\text{т}},\ d_{\text{т}},\ n_{\text{т}}) + \Delta p_{\text{н}}\,(l_{\text{н}},\ d_{\text{н}},\ n_{\text{н}}) = p^0_{cj} - p^k_{cj} \end{aligned} \qquad \text{(VI, 75)}$$

Вся остальная процедура синтеза ТС аналогична описанной выше.

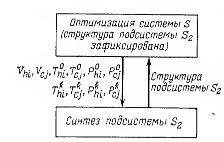


Рис. 45. Схема процедуры синтеза теплообменной системы как части ХТС произвольной структуры.

Перейдем теперь к основной задаче. Воспользуемся принципом закрепления (см. с. 170) и закрепим температуры (VI, 55), (VI, 56), давления (VI, 73) и расходы (VI, 74). Благодаря тому, что взаимное влияние подсистем S_1, S_2 будет ликвидировано, оптимизацию подсистемы S_1 и синтез подсистемы S_2 (первый уровень процедуры) можно провести отдельно. В результате получим какое-то значение критерия оптимизации F для всех схемы, которое будет функцией переменных (VI, 55), (VI, 56), (VI, 73), (VI, 74):

$$F = F\left(T^0_{hi},\ T^0_{cj},\ T^k_{hi},\ T^k_{cj},\ p^0_{hi},\ p^0_{cj},\ p^k_{hi},\ p^k_{cj},\ V_{hi},\ V_{cj}\right) \qquad \text{(VI, 76)}$$

Процедуре 2-го уровня будет соответствовать задача оптимизации функции непрерывных переменных (VI, 76). Этот подход имеет один недостаток. Поскольку при синтезе подсистемы приходится решать комбинаторную задачу, относительно гладкости функции (VI, 76) ничего сказать нельзя. Во всяком случае, трудно предполагать существование во всей области определения не только вторых, но и первых производных данной функции. Это будет препятствовать применению наиболее эффективных поисковых методов — квазиньютоновских; т. е. для оптимизации функции (VI, 76) можно будет применять только методы нулевого порядка.

Рассмотрим теперь другой подход. Он также будет двухуровневым и основывается на принципе закрепления. Пусть опять закреплены переменные (VI,55), (VI,56), (VI,73), (VI,74). Проведем синтез подсистемы S_2 (первый уровень). На второй уровень вынесем задачу оптимизации всей системы S. При этом в подсистеме S_2 будут оптимизироваться только технологические параметры — длины, диаметры и число трубок, расходы пара в нагревателе и охлаждающей воды в холодильнике, а в подсистеме S_1 — все варьируемые параметры. После решения этой задачи получим новые значения переменных (VI, 55), (VI, 56), (VI, 73), (VI, 74) на входе и выходе ТС (подсистемы S_2) и можно опять переходить к первому уровню — решению задачи синтеза ТС, и т. д. (рис. 45). Преимущество этого подхода перед предыдущим состоит в том, что критерий оптимизации в данном случае является достаточно гладкой функцией, для минимизации которой можно использовать квазиньютоновские методы 1-го порядка. Легко видеть, что описанная двухуровневая процедура применима с небольшими изменениями и в случае, когда S_2 — произвольная подсистема.

Общий подход, развитый выше, проиллюстрируем на примере, когда S_1 является системой ректификационных колонн, а S_2 — теплообменной системой. В каждой ректификационной колонне

поток, отходящий из верхней части должен быть охлажден в конденсаторе-холодильнике (будем далее называть его холодильником), а к возвращаемой части потока, отбираемого из низа, должно быть подведено тепло; обеспечивающее испарение, для чего предусмотрен нагреватель (кипятильник). С точки зрения задачи синтеза ТС поток из верха колонны будет «горячим», а возвращаемая часть потока из низа колонны «холодным». В схемах разделения часто используются только внешние источники энергии (каждый горячий поток охлаждается своим холодильником, а каждый «холодный» нагревается своим нагревателем), такие ТС не содержат рекуперативных теплообменников. В то же время тепло «горячих» потоков может быть использовано для испарения «холодных» потоков. На рис. 46 приведена схема разделения пятикомпонентной смеси *ABCDE*, в которой горячий поток, отбираемый из верха колонны *3*, используется в рекуперативном теплообменнике для подвода тепла к возвращаемой части потока, отбираемого из низа колонны *4*. В схеме используются внутренние источники энергии.

Пусть структура схемы разделения задана и задача состоит в выборе структуры ТС и одновременной оптимизации всей схемы в целом. Описанная выше процедура полностью применима и в этой задаче. Включим в совокупность S_h горячих потоков все потоки из верхних частей колонн, а возвращаемые части потоков, отобранные из их нижней частей, — в совокупность S_c холодных потоков. Если между какими-либо холодным и горячим потоками теплообмен невозможен, это можно учесть при решении задачи синтеза ТС. Закрепим переменные (VI, 55), (VI, 56), (VI, 73), (VI, 74) на входе и выходе ТС и проведем синтез ТС, используя один из описанных методов с учетом ограничений (VI, 75). После того как структура ТС будет найдена, проведем оптимизацию всей схемы, используя в качестве управляющих переменных ректификационных колонн флегмовые числа, числа тарелок и т. д., а теплообменных систем — числа, диаметры и длины трубок. В результате получим новые значения всех или некоторых переменных (VI,55), (VI,56), (VI,73), (VI,74). Закрепим их и вновь проведем решение задачи синтеза ТС и т. д. Синтез систем ректификации с рекуперацией тепла см. в Приложении 3.

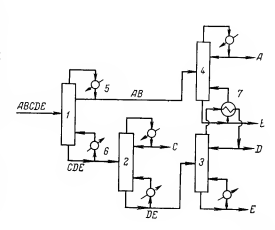


Рис. 46. Схема разделения смеси *ABCDE*, в которой используются внутренние источники энергии:
1—*4* — ректификационные колонны; *5* — холодильник; *6* — нагреватель; *7* — рекуперативный теплообменник.

Синтез многослойного адиабатического реактора

Пусть имеется многослойный каталитический реактор (рис. 47), в который поступает исходный поток (температура T^0, расход G^0). Будем считать, что в слоях идет экзотермическая реакция. Обычно исходный поток используется для охлаждения промежуточных потоков между слоями. Пусть для охлаждения используется только процесс теплообмена (без смешения). Известная задача [145] состоит в определении таких времен контакта τ_i ($i = \overline{1, N}$) и температур входных потоков в слои, чтобы некоторый критерий F (будем считать, что это затраты) принял минимальное значение. Эту задачу интересно решить совместно с задачей выбора структуры ТС, обеспечивающей необходимый температурный режим в слоях.

Воспользуемся общей схемой, изложенной в предыдущем разделе. В данном случае, в качестве подсистемы S_1 выступают N слоев катализатора, а в качестве S_2 — теплообменная система. Рассмотрим вначале случай, когда стоимость ТС мала по сравнению со стоимостью слоев катализатора. В этом случае вначале может быть решена задача оптимизации многослойного реактора без ТС, причем предполагается, что входные температуры $T_{\text{вх}}^{(i)}$ потоков в слои являются независимыми поисковыми переменными, а выходные $T_{\text{вых}}^{(i)}$ — свободными. После решения задачи оптимизации для каждого слоя будут известны входные и выходные температуры потоков. Поскольку реакция экзотермическая, $T_{\text{вых}}^{(i)} > T_{\text{вх}}^{(i)}$. Таким образом, температуру потока между i-тым и $(i+1)$-м слоем необходимо понизить от вели-

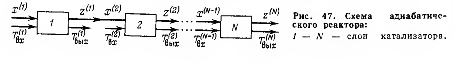


Рис. 47. Схема адиабатического реактора:
1 — N — слои катализатора.

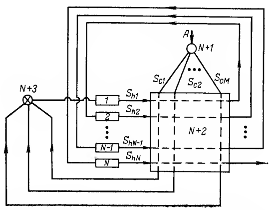


Рис. 48. Глобальная схема многополочного адиабатического реактора:
1 — N — слои катализатора; $N+1$ — делитель потока; $N+2$ — теплообменная система; $N+3$ — смеситель; A — исходный поток.

чины $T_{\text{вых}}^{(i)}$ до величины $T_{\text{вх}}^{(i+1)}$. Пользуясь терминологией задачи синтеза ТС, можем сказать, что имеется один исходный холодный поток и N горячих (выходных потоков из слоев), причем $N - 1$ горячих потоков (соответствующих промежуточным потокам между слоями) необходимо охладить от температуры $T_{\text{вых}}^{(1)}, \ldots, T_{\text{вых}}^{(N-1)}$ до температур $T_{\text{вх}}^{(2)}, \ldots, T_{\text{вх}}^{(N)}$ соответственно. Глобальная схема синтеза многополочного реактора совместно с ТС [146] приведена на рис. 48.

Чтобы увеличить возможности ТС, исходный поток A делится на M потоков в делителе, см. блок $(N + 1)$. Пройдя через ТС, эти потоки смешиваются в смесителе $(N + 3)$, откуда поток подается в 1-й слой. Используя одну из процедур синтеза ТС, можно найти ее структуру. Рассмотрим теперь случаи, когда стоимость ТС сравнима со стоимостью слоев катализатора. В этом случае на первом уровне процедуры синтеза необходимо будет найти оптимальную структуру ТС при фиксированных значениях входных и выходных температур слоев. На верхнем же уровне необходимо будет совместно оптимизировать систему слоев и ТС. При этом поисковыми переменными будут длины слоев катализатора, поверхности теплообмена и структурные параметры, соответствующие делителю $(N + 1)$. Поскольку в этом случае все переменные непрерывны, для оптимизации системы могут быть использованы поисковые методы (см. гл. III, IV).

ЛИТЕРАТУРА

1. *Островский Г. М., Волин Ю. М.* Методы оптимизации сложиых химико-технологических схем. М., Химия, 1970. 328 с.
2. *Byron S. G., Paul R. B., Eldon P. H.* — Ind. Eng. Chem. Proc. Des. Develop., 1970, v. 9, № 4, p. 581.
3. *Островский Г. М., Волин Ю. М.* Моделироваиие сложиых химико-технологических схем. М., Химия, 1975. 311 с.
4. *Katz S.* — Ann. N. J. Acad. Sci., 1960, v. 84, p. 441.
5. *Арис Р.* Оптимальное проектироваиие химических реакторов/Пер. с айгл. М., Издатинлит, 1963. 238 с.
6. *Fan L. T.* The Continuous Maximum Principle. New York, John Wiley and Sons, 1966. 410 p.
7. *Фан Л. Ц., Вань Ч. С.* Дискретиый прииции максимума/Пер. с англ. под ред. А. И. Пропоя. М., Мир, 1967. 180 с.
8. *Spendley W., Hext G. R., Himsworth F. R.* — Technometrics, 1962, v. 4, № 4, p. 441.
9. *Box M. J.* — Computer J., 1965, v. 8, № 1, p. 442.
10. *Powell M. J. D.* — Computer J., 1964, v. 7, № 2, p. 155—162.
11. *Островский Г. М., Бережинский Т. А., Беляева А. Р.* Алгоритмы оптимизации химико-технологических процессов. М., Химия, 1978. 296 с.
12. *Растригин Л. А.* Статистические методы поиска. М., Наука, 1968. 320 с.
13. *Растригин Л. А.* Системы экстремального управлеиия. М., Наука, 1974. 630 с.
14. *Westerberg A. W., Hutchison H. P., Motard R. L. e. a.* Process Flowsheeting. Cambridge, Cambridge University Press, 1979. 251 p.
15. *Tavast R., Roose A.* — In: Proceedings of the Symposium «Computers in the Design and Evectoon of Chemical Plants», 1975, v. 1, p. 443—450.

16. *Umeda T.*, *Nishio M.* — Ind. Eng. Chem. Proc. Des. Develop., 1972, v. 11, № 2, p. 153—161.
17. *Kevorkian A. K.*, *Snoek J.* — In: Dekomposition of Large Scale Problems, ed. by D. M. Himmelblau. Amsterdam, North Holland, 1973, p. 467—489.
18. *Кафаров В. В.*, *Перов В. Л.*, *Мешалкин В. П.* Принципы математического моделирования химико-технологических систем. М., Химия, 1974. 344 с.
19. *Motard R. L.*, *Shacham M.* — AJChE J., 1975, v. 21, № 3, p. 417—436.
20. *Стренг Г.* Линейная алгебра и ее применения/Пер. с англ. под ред. Г. И. Марчука. М., Мир, 1980. 454 с.
21. *Brown K. M.* — In: Numerical Solution at Systems at Nonlinear Algebraical Equations, ed. by J. D. Byr. New York, Academic Press, 1973, p. 281—340.
22. *Orbach O.*, *Crove C. M.* — Can. J. Chem. Eng., 1971, v. 49, № 8, p. 509—513.
23. *Crow C. M.*, *Nishio M.* — AJChE J., 1975, v. 21, № 3, p. 528—533.
24. *Broyden C. G.* — Mathematic. Comp., 1965, v. 19, № 91, p. 577—593.
25. *Островский Г. М.*, *Бережинский Т. А.*, *Алиев К. А.* — Азерб. хим. ж., 1979, № 2, с. 26—28.
26. *Островский Г. М.*, *Волин Ю. М.*, *Ханзель К.* Расчет стационарных режимов химико-технологических схем, вып. 2/104/. М., НИИТЭхим, 1981. 64 с.
27. *Greenstadt J.* — Mathematic. Comp., 1970, v. 24, № 109, p. 1—22.
28. *Dennis J. E.* — Schnabel SIAM Revue, 1979, v. 21, № 4, p. 443—459.
29. *Gay D. M.* — SIAM J. Numer. Analysis, 1979, v. 16, № 4, p. 623—630.
30. *Metcalfe S. K.*, *Perkins J. D.* — Trans. Inst. Chem. Eng., 1978, v. 56, № 2, p. 210.
31. *Pearson J. D.* — Computer J., 1969, v. 12, № 2, p. 171—178.
32. *Murtagh B. A.*, *Sergent R. W. H.* — In: Optimization, ed. by R. Fletcher. London — New York, Academic Press, 1969, p. 215—246.
33. *Adachi N. J.* — Optim. Theory Applics, 1971, v. 7, № 6, p. 391—410.
34. *Barnes J.* — Computer J., 1965, v. 8, № 1, p. 66—72.
35. *Ортега Дж.*, *Рейнболдт В.* Итерационные методы решения систем уравнений со многими неизвестными/Пер. с англ. под ред. И. В. Коновалова. М., Мир, 1975. 558 с.
36. *Тьюрсон Р.* Разреженные матрицы/Пер. с англ. под ред. Х. К. Икрамова. М., Мир, 1977. 187 с.
37. *Wegstein J. H.* — Comm. ACM, 1958, v. 1, № 6, p. 9.
38. *Williams T. J.*, *Otto R. E.* — AIEE Trans., 1960, v. 79, № 11, p. 458—473.
39. *Ray W. H.*, *Szekely J.* Process Optimization with Applications in Metallurgy and Chemical Engineering. New York, John Wiley and Sons, 1973. 371 p.
40. Справочник нефтехимика, т. 2. М., Химия, 1978.
41. *Маслянский Г. Н.*, *Бурсиан Н. Р.*, *Баркан С. А.* — ЖПХ, 1966, т. 39, № 3, с. 342—351.
42. *Бурсиан Н. Р.*, *Маслянский Г. Н.* — Хим. пром., 1961, № 3, с. 152—158.
43. *Анисимов И. В.*, *Бодров В. И.*, *Покровский В. Б.* Математическое моделирование и оптимизация ректификационных установок. М., Химия, 1975. 216 с.
44. *Александров И. А.* Ректификационные и абсорбционные аппараты. М., Химия, 1978. 280 с.
45. *Каневец Г. Е.* Обобщенные методы расчета теплообменников. Киев, Наукова думка, 1979. 352 с.
46. *Hartman K.* (ed.), Modellierung und Optimierung Verfahrenstechnischer Systeme. Berlin, Akademie Verlag, 1978. 335 S.
47. *Ханзель К.*, *Волин Ю. М.*, *Островский Г. М.* Методы структурного анализа в задачах исследования химико-технологических схем, вып. 4/89/. М., НИИТЭхим, 1980. 58 с.
48. *Мотыль Д. Н.*, *Волин Ю. М.*, *Островский Г. М.* — ТОХТ, 1981, т. 15, № 2, с. 232—245.
49. *Shubert L. K.* — Mathematic. Comp., 1970, v. 24, № 109, p. 27—30.
50. *Островский Г. М.*, *Бережинский Т. А.*, *Слинько М. Г.* и др. — ДАН СССР, 1980, т. 253, № 2, с. 425—427.
51. *Соколинский Ю. А.* Канд. дис. М., МИХМ, 1963.
52. *Larson A. T.*, *Black C. A.* — J. Am. Chem. Soc., 1925, April, p. 1015.

53. *Раскин А. Я., Соколинский Ю. А., Ягнятинский Б. В.* и др. — Труды ГИАП. М., 1975, вып. 37, с. 5—22.
54. *Мотыль Д. Н.* Канд. дисс., М., ГИАП, 1980.
55. *Box H. J.* — Computer J., 1966, v. 9, № 1, p. 67—77.
56. *Мальцев А. И.* Основы линейной алгебры. М., Наука, 1970. 400 с.
57. *Пшеничный Б. Н., Данилин Ю. М.* Численные методы в экстремальных задачах. М., Наука, 1975. 319 с.
58. *Fletcher R., Reeves C. M.* — Computer J., 1964, v. 7, № 2, p. 149—154.
59. *Fletcher R.* — In: Numerical Methods for Unconstrained Optimization, ed. by Murray. London — New York, Academic Press, 1972, p. 73—86.
60. *Boland W. R., Kamgnia E. R., Kowalik J. S.* — J. Optim. Theory Applics, 1979, v. 27, № 2, p. 221—229.
61. *Kowalik J. S., Kamgnia E. R., Boland W. R.* — J. of Mathematic. Anal. a. Applications, 1979, v. 67, № 2, p. 476—482.
62. *Powell M. J. D.* — In: Nonlinear Programming, ed. by J. B. Rosen, O. L. Mangasarian, K. Ritter. New York, Academic Press, 1970, p. 31—43.
63. *Fletcher R., Powell M. J. D.* — Computer J., 1963, v. 6, № 2, p. 163—168.
64. *Broyden C. G.* — J. Inst. Maths. Appl., 1970, v. 6, № 3, p. 222—237.
65. *Fletcher R.* — Computer J., 1970, v. 13, № 3, p. 317—322.
66. *Huang H. Y.* — J. Optim. Theory Applics, 1970, v. 5, № 6, p. 405—423.
67. *Broyden C. G.* — Mathematic. Comp., 1967, v. 21, № 99, p. 368—381.
68. *Broyden C. G.* — J. Inst. Maths Applics, 1970, v. 6, № 1, p. 76—90.
69. *Oren S. S., Luenberger D.* — Management Sci., 1974, v. 20, № 5, p. 845—862.
70. *Oren S. S.* — Management Sci., 1974, v. 20, № 5, p. 863—874.
71. *Oren S. S.* — Mathematic. Progr., 1974, v. 7, № 3, p. 351—367.
72. *Dixon L. C. W.* — J. Optim. Theory Appl., 1972, v. 10, № 1, p. 34—40.
73. *Островский Г. М., Волин Ю. М.* Методы оптимизации химических реакторов. М., Химия, 1967. 248 с.
74. *Morrison D.* — SIAM J. Numer. Anal., 1968, v. 5, № 1, p. 83—88.
75. *Кирин Н. Е.* Вычислительные методы теории оптимального управления. Л., Изд-во Ленинградского ун-та, 1968. 144 с.
76. *Дей М. М.* Нормированные линейные пространства/Пер. с англ. М., Издатинлит, 1961. 232 с.
77. *Островский Г. М.* Докт. дис. М., НИФХИ им. Л. Я. Карпова, 1966.
78. *Kowalik J., Osborne M. R., Ryan D. M.* — Operat. Res., 1969, v. 17, № 6, p. 973—983.
79. *Hestenes M. R.* — J. Optim. Theory Appl., 1969, v. 4, № 5, p. 303—321.
80. *Иоффе А. Д., Тихомиров В. М.* Теория экстремальных задач. М., Наука, 1974. 480 с.
81. *Шаманский В. Е.* Методы численного решения краевых задач на ЭЦВМ, ч. 2. Киев, Наукова думка, 1966. 244 с.
82. *Гандельсман Т. А., Бесков В. С., Козлов В. П.* и др. — Хим. пром., 1971, № 10, с. 753—757.
83. *Гандельсман Т. А., Козлов В. П., Родов А. Б.* и др. — В кн.: Математическое моделирование сложных химико-технологических систем. Доклады первой Всесоюз. конф. по математическому моделированию сложных химико-технологических систем (СХТС-1). Ереван, Изд-во Ереванского Гос. ун-та, 1975, с. 339—343.
84. *Боресков Г. К., Буянов Р. А., Иванов А. А.* — Кинетика и катализ, 1967, т. 8, № 1, с. 153.
85. *Слинько М. Г., Мулер А. Л.* — Кинетика и катализ, 1961, т. 2, № 2, с. 467.
86. *Боресков Г. К., Слинько М. Г., Бесков В. С.* — Хим. пром., 1968, № 3, с. 173—178.
87. *Островский Г. М., Бережинский Т. А., Волин Ю. М.* и др. — В кн.: Математические методы в химии. Материалы 3 Всесоюз. конф. «Математические методы в химии» (ММХ-3). М., ЦНИИТЭнефтехим, 1980, с. 39—46.
88. *Деннис Дж.* Математическое программирование и электрические цепи/Пер. с англ. М., Издатинлит, 1961. 215 с.
89. *Goldfarb D.* — SIAM J. Appl. Math., 1969, v. 17, p. 739—762.
90. *Гилл Ф., Мюррей У.* Численные методы условной оптимизации/Пер. с англ. под ред. А. А. Петрова. М., Мир, 1977. 296 с.

91. *Gill P., Murray W.* — J. Inst. Math. Appl., 1972, v. 9, p. 91—108.
92. *Abadie J., Carpentier J.* — In: Optimization, ed. by R. Fletcher. London, Academic Press, 1969, p. 49—64.
93. *Wolfe P.* — In: Recent Advances in Mathematical Programming, ed. by R. L. Graves and P. Wolfe. New York, McGraw-Hill, 1963, p. 67—86.
94. *Фельдбаум А. А.* Основы теории оптимальных автоматических систем. М., Физматгиз, 1963. 552 с.
95. *Островский Г. М.* — Изв. АН СССР. Сер. техн. кибернетика, 1964, № 5, с. 64—72.
96. *Волин Ю. М., Островский Г. М.* — Изв. АН СССР. Сер. техн. кибернетика, 1965, № 5, с. 137—142.
97. *Бережинский Т. А., Островский Г. М., Волин Ю. М.* — ТОХТ, 1967, т. 1, № 5, с. 699—710.
98. *Химмельблау Д.* Прикладное нелинейное программирование/Пер. с англ. под ред. М. Л. Быховского. М., Мир, 1975. 534 с.
99. *Подвальный С. Л.* и др. — В кн.: Автоматизация химических производств, 1976, № 6, с. 16—20.
100. *Подвальный С. Л.* Моделирование промышленных процессов полимеризации. М., Химия, 1979. 256 с.
101. *Островский Г. М., Михайлова Е. М., Бережинский Т. А.* и др. — В кн.: Тезисы докладов Второй Всесоюз. конф. «Математическое моделирование сложных химико-технологических систем» (СХТС-2). Новомосковск, 1979, с. 82—83.
102. *Кафаров В. В., Золотарев В. В., Михайлов Г. В.* и др. — Хим. пром., 1981, № 9, с. 548.
103. *Trefny F.* — Chem. Ind. Techn., 1965, v. 10, № 2, p. 191.
104. *Маньковский О. Н., Толчинский А. Р., Александров М. В.* Теплообменная аппаратура химических производств. Л., Химия, 1976. 328 с.
105. *Витков В. С.* Канд. дис. М., МХТИ им. Д. И. Менделеева, 1978.
106. *Михайлов Г. В., Витков В. С.* — В кн.: Математическое моделирование технологических процессов в производствах минеральных удобрений и серной кислоты. Труды НИУИФ, 1980. вып. 236, с. 5.
107. *Sargent R. W. U.* A Review of Optimization Methods for Nonlinear Problems, presented at Symposium on Computer Applications to Chemical Engineering Process Design and Simulation, ACS Annual Meeting. Wathington, 1979.
108. *Волин Ю. М., Островский Г. М.* — Автоматика и телемеханика, 1966, т. 27, № 12, с. 29—36.
109. *Волин Ю. М., Островский Г. М.* — Изв. АН СССР. Сер. техн. кибернетика, 1980, № 6, с. 195 (деп.).
110. *Островский Г. М.* — В кн.: Кибернетику на службу коммунизму/Под. ред. А. И. Берга. М.-Л., Энергия, 1967, с. 302—308.
111. *Umeda T., Shindo A., Tazaki E.* — Ind. Eng. Chem. Proc. Dis. Develop., 1972, v. 11, № 7, p. 1—8.
112. *Lasdon L. S., Schoeffler J.* — Proc. IAAC, 1965, Troy, New York.
113. *Toint Ph. L.* — Mathematic. Comp., 1967, v. 21, № 100, p. 954—961.
114. *Shanno D. F.* — Mathematic. Comp., 1980, v. 34, № 150, p. 499—514.
115. *Островский Г. М., Бережинский Т. А., Беляева А. Р.* и др. — ТОХТ, 1980, т. 14, № 3, с. 423—430.
116. *Hendry J. E., Rudd D. F., Seader J. D.* — AIChE J., 1973, v. 19, № 1, p. 1—15.
117. *Ponton J. W., Donaldsen R. A.* — Chem. Eng. Sci., 1974, v. 29, № 12, p. 2375—2377.
118. *Umeda T., Ichikawa A.* An Unifieci Approach to Processing System Design, Proc. USA — Japan Seminar «The Application Process Systems Engineering to Chemical Technology Assessment», Kyoto, Tokyo Institute of Technology. Tokyo, 1975. 129 p.
119. *Кафаров В. В., Мешалкин В. П., Перов В. Л.* Математические основы автоматизированного проектирования химических производств. М., Мир, 1979. 318 с.
120. *Корбут А. А., Финкельштейн Ю. Ю.* Дискретное программирование, М., Наука, 1969. 366 с.

121. *Рафаэл Б.* Думающий компьютер/Пер. с англ. под ред. В. Л. Стефанюка. М., Мир, 1979. 405 с.
122. *Fan L. T., Chen H. T., Aldis D.* — In: Proceedings of the Symposium «Computers in the Design and Erection of Chemical Plant», v. 1, Karlovy Vary (Chechoslovakia), 1975, p. 279—291.
123. *Касти Дж., Калаба Р.* Методы погружения в прикладной математике/Пер. с англ. под ред. С. П. Чеботарева. М., Мир, 1976. 220 с.
124. *Беллман Р., Дрейфус С.* Прикладные задачи динамического программирования/Пер. с англ. под ред. А. А. Первозванского. М., Наука, 1965. 457 с.
125. *Cruhn G., Hartman K., Kötchen P. e. a.* Systemverfahrenstechnik, Bd. 2. Leipzig, VEB Deutcher Verlag für Grundstoffindustrie, 1977. 186 S.
126. *Pho T. K., Lapidus L.* — AIChE J., 1969, v. 19, № 6, p. 1182—1189.
127. *Кофман А.* Введение в прикладную комбинаторику/Пер. с франц. под ред. Б. А. Севастьянова. М., Мир, 1975. 479 с.
128. *Слинько М. Г.* Моделирование химических реакторов. Новосибирск, Наука, 1968. 95 с.
129. *Rodrigo R. F., Seader J. D.* — AIChE J., 1975, v. 21, № 5, p. 885—894.
130. *Umeda T., Hirai A., Ichikawa A.* — Chem. Eng. Sci., 1972, v. 27, p. 795—804.
131. *Shah J. V., Westerberg A. W.* — AIChE J., 1977, v. 23, № 3, p. 378—380.
132. *Островский Г. М., Шевченко А. Л.* — ТОХТ, 1979, т. 13, № 3, с. 428—435.
133. *Ostrovskij G. M., Shevchenko A. L.* — Chem. Eng. Sci., 1979, v. 34, p. 1243—1245.
134. *Ostrovskij G. M., Shevchenko A. L., Bereshinskij T. A.* — Wissenschaflicke Zeitschrift TH Leuna — Merseburg, 1981, Bd. 23, № 2, S. 244—251.
135. *Hwa C. S.* — AIChE — Int. Chem. E. Symp. Ser., 1965, № 4, p. 101—105.
136. *Lee K. E., Masso A. H., Rudd D. F.* — Ind. Eng. Chem. Fundamentl. 1970, v. 9, № 1, p. 48.
137. *Nishimura U., Niraizumi Y., Kunii D.* Soc. Chem. Eng. Japan Autumnal Mtg., 1968, E-214 and E-215.
138. *Kobayshi S., Umeda T., Ichikawa A.* — Chem. Eng. Sci., 1971, v. 26, № 7, p. 1367—1380.
139. *Кафаров В. В., Мешалкин В. П.* — ДАН СССР, 1979, т. 246, № 6, с. 1435—1439.
140. *Rockstroch L., Hartman K.* — Chem. Techn., 1975, Bd. 27, № 8, S. 439—442.
141. *Ивахненко В. И., Островский Г. М.* — ТОХТ, 1982, т. 16, № 3, с. 348—354.
142. *Островский Г. М., Шевченко А. Л., Сальникова Г. Н.* и др. — Хим. технол., 1978, № 4, с. 28—32.
143. *Rathor R. N. S., Powers G. J.* — Ind. and Eng. Chem. Proc. Des. Develop., 1975, v. 14, № 2, p. 175—187.
144. *Grossman I. E., Sargent R. W. H.* — Comp. Chem. Eng., 1978, v. 2, № 1, p. 1—7.
145. *Боресков Г. К., Слинько М. Г.* — ЖПХ, 1943, т. 16, № 4, с. 377.
146. *Островский Г. М., Слинько М. Г., Шевченко А. Л.* и др. ДАН СССР, 1979, т. 247, № 6, с. 1433—1436.
147. *Фадеев Д. К., Фадеева В. Н.* Вычислительные методы линейной алгебры. М., Физматгиз, 1960. 656 с.
148. *Фихтенгольц Г. М.* Курс дифференциального и интегрального исчисления. 5-е изд., т. 1. М., Наука, 1962. 607 с.
149. *Карманов В. Г.* Математическое программирование. М., Наука, 1980. 256 с.
150. *Гантмахер Ф. Р.* Теория матриц. М., Наука. 1967. 575 с.
151. *Penrose R.* — Proc. Camb. Phil. Soc., 1955, v. 51, № 2, p. 406—413.
152. *Пшеничный Б. Н.* — ДАН СССР, 1977, т. 235, № 5, с. 1026—1029.
153. *Левитин Е. С., Поляк Б. Т.* — Журн. вычислит. матем. и матем. физики, 1966, т. 6, № 5, с. 787—823.
154. *Dennis J. E., More J. J.* — Maths. Comp., 1974, v. 28, № 126, p. 549—560.
155. *Broyden C. G., Dennis J. E., More J. J.* — J. Inst. Maths. Appl., 1973, v. 12, № 3, p. 223—245.
156. *Rosenbrock H. H.* — Computer J., 1960, v. 3, № 3, p. 175—184.
157. *Parkinson J. M., Hutchison D.* — In: Numerical Methods for Nonlinear Optimization. ed. by F. E. Lootsma. London — New York, Academic Press, 1972, p. 99—113.

158. *Brent R. P.* Algorithms for Minimization without Derivatives. Englewood Cliffs — New Jersey, Prentice Hall Inc., 1973. 195 p.
159. *Dixon L. C. W.* — J. Inst. Maths. Appl., 1973, v. 11, № 3, p. 317—328.
160. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации/Пер. с англ. под ред. Е. Г. Гольштейна. М., Мир, 1972. 240 с.
161. *Dixon L. C. W.* — In: Numerical Methods for Nonlinear Optimization, ed by F. A. Lootsma. London — New York, Academic Press, 1972, p. 149—170.
162. *Biggs M. C.* — J. Inst. Maths. Appl., 1971, v. 8, № 3, p. 315—327.
163. *Shanno D. F.* — Mathematic. Comp., 1970, v. 24, № 111, p. 647—656.
164. *Pierre D. A., Lowe M. J.* Mathematical Programming via Augmented Lagrangians. London, Addison — Wesley Publishing Company, 1975. 436 p.
165. *Sophos A., Rotstein E., Stephanopoulos G.* — Chem. Eng. Sci., 1980, v. 35, № 12-D, p. 2415—2426.
166. *Morari M., Faith D. C.* — AIChE J., 1980, v. 26, № 6, p. 916—928.

ПРИЛОЖЕНИЕ 1. МНОГОКРИТЕРИАЛЬНАЯ ОПТИМИЗАЦИЯ

Предмет так называемой многокритериальной оптимизации составляют задачи типа

$$f,\ \varphi_1,\ \varphi_2, \ldots,\ \to \min$$

В этом случае поставленная цель характеризуется не одной функцией f, а несколькими $f, \varphi_1, \varphi_2, \ldots$, и задача оптимизации заключается в поиске условного (или безусловного) минимума всех критериев. Отметим, что критерии $f, \varphi_1, \varphi_2, \ldots$, вообще говоря, противоречивы в том смысле, что точки в пространстве варьируемых переменных, представляющие собой решения задач оптимизации по каждому из заданных критериев, в общем случае различны и, следовательно, ни одна из них не может служить решением исходной многокритериальной задачи.

Рассмотрим пространство E^m значений критериев (m — число рассматриваемых критериев). Каждой точке $x \in \Omega \subset E^h$ пространства варьируемых переменных соответствует точка $[f(x), \varphi_1(x), \varphi_2(x), \ldots] \in E^m$. Образом множества Ω при этом отображении E^n в E^m является область Λ достижимых значений векторного критерия $(f, \varphi_1, \varphi_2, \ldots)$ (см. рис. 14). Внутренние точки Λ не могут содержать решения задачи, так как в этом случае в окрестности рассматриваемой точки $z(x)$ всегда найдется точка из Λ с меньшим значением критериев (в заштрихованной части множества Λ).

Назовем точку $\bar{x} \in \Omega$ неулучшаемой, если ни при каком $\bar{x} \in \Omega$ не выполняются неравенства

$$f(\bar{x}) \leqslant f(x) \qquad \varphi_1(\bar{x}) \leqslant \varphi_1(x) \qquad \varphi_2(\bar{x}) \leqslant \varphi_2(x),\ \ldots$$

причем по крайней мере одно из них должно быть строгим. Ясно, что образом множества неулучшаемых точек является часть границы Λ (на рис. 14 обозначена жирной линией AB), ближайшая к так называемой, утопической точке $C = (\varphi_{\min}, f_{\min})$. Множество неулучшаемых точек (так называемое множество Парето) и следует формально считать решением исходной многокритериальной задачи. Для получения единственного решения данной задачи необходима дополнительная информация, позволяющая произвести свертку исходных критериев в один, так называемый «глобальный» критерий (т. е. сформировать скалярную функцию, зависящую от исходных критериев), и свести тем самым исходную задачу к традиционной однокритериальной задаче. Таким образом, решение многокритериальной задачи содержит два этапа: построение множества Парето и выбор «предпочтительного» решения, являющегося элементом этого множества, на основе некоторого «глобального» критерия.

Одним из способов построения множества Парето является, так называемый «параметрический» метод [165], в котором осуществляется свертка критериев во взвешенную сумму

$$F(f,\ \varphi_1,\ \varphi_2, \ldots) = \sum_{i=1}^{m-1} \omega_i \varphi_i + \omega_0 f \qquad \sum_{i=0}^{m-1} \omega_i = 1 \qquad \omega_i \geqslant 0$$

с последующим нахождением ее минимума на Ω. Элементы множества Парето могут быть найдены при изменении ω_i. Все элементы такого множества могут быть найдены этим способом лишь для выпуклого Λ. При наличии «зазоров» между границей Λ и опорными к Λ плоскостями (невыпуклый случай) могут быть определены лишь элементы некоторого подмножества множества Парето (на рис. 14 элементы множества Парето, которым соответствуют точки дуги AB, расположенные между точками 1 и 2, не могут быть найдены параметрическим методом). В связи с этим выбор взвешенной суммы в качестве «глобального» критерия для невыпуклых задач является проблематичным, так как при выборе «предпочтительного» решения элементы множества Парето, образы которых расположены в «зазоре» Λ, не рассматриваются.

Более универсальными методами построения множества Парето, пригодными и в отсутствие выпуклости Λ, являются методы последовательной безусловной минимизации («уровней», штрафных функций, модифицированной функции Лагранжа), которые применяются к решению задачи минимизации одного из критериев, например f, с ограничениями, определяемыми функциями φ_i всех прочих критериев.

Составная функция конкретного метода последовательной безусловной минимизации, очевидно, представляет собой некоторую свертку исходных критериев, зависящую от параметров. Из геометрических соображений, приведенных в гл. IV при изложении методов последовательной безусловной минимизации, следует, что минимум функции свертки при фиксированных значениях параметров определяет элемент множества Парето, поскольку изоповерхность функции свертки, рассматриваемая в пространстве критериев и отвечающая минимальному ее значению, касается границы Λ. Таким образом, для задачи минимизации f при ограничениях $\varphi_i = 0$ элементом множества Парето является не только x^* — решение задачи, но и точки x, определяемые в процессе решения и соответствующие минимуму составной функции (свертки) при фиксированных значениях ее параметров. Для получения достаточно полного представления о множестве Парето необходимо, как и при использовании параметрического метода, выполнить неоднократное решение задачи: $\min f$ при $\varphi_i = a_i$ для различных a_i, $a_{i_{\min}} \leqslant a_i \leqslant a_{i_{\max}}$ (на рис. 14, $\varphi_A^* \leqslant a \leqslant \varphi_B$).

На втором этапе решения многокритериальной задачи (выбор «наилучшей» точки из множества Парето) наиболее целесообразно воспользоваться ее близостью к утопической точке C в смысле некоторой меры, например, квадратичной метрики $F(f, \varphi) = (f - f_{\min})^2 + (\varphi - \varphi_{\min})^2$ или $\overline{F}(f, \varphi) = \max \{|f - f_{\min}|, |\varphi - \varphi_{\min}|\}$ определяющей необходимую свертку, т. е. «глобальный» критерий. В первом случае решению соответствует точка P касания с множеством Λ сферы с центром в C и наименьшим радиусом, во втором — точка $Q \in \Lambda$, имеющая минимальное значение наибольшего из отклонений от утопической точки.

ПРИЛОЖЕНИЕ 2. КУБИЧЕСКАЯ ИНТЕРПОЛЯЦИЯ

Пусть на отрезке $[0,1]$ задан полином $f(x)$ третьей степени $f(x) = ax^3 + bx^2 + cx + d$ и $g(x)$ — его градиент: $g(x) = 3ax^2 + 2bx + c$. Определим положение минимума $\alpha \in [0, 1]$ функции $f(x)$ при заданных значениях f и g на концах отрезка: $f_0 = f(0)$; $g_0 = g(0)$; $f_1 = f(1)$; $g_1 = g(1)$.

В точке минимума $x = \alpha$ должно быть $g(\alpha) = 0$, $g'(\alpha) > 0$. Отсюда

$$\alpha = \frac{-b + \sqrt{b^2 - 3ac}}{3a} \qquad (b^2 - 3ac) > 0$$

Условия на концах отрезка дают:

$$f_0 = d; \quad g_0 = c; \quad f_1 = a + b + c + d; \quad g_1 = 3a + 2b + c$$

Обычно в алгоритмах минимизации добиваются $g_0 < 0$, т. е. $c < 0$, что мы и будем предполагать. Последние два условия (с учетом двух первых) можно рассматривать как систему линейных уравнений относительно a и b. Находим

$$a = 2(f_0 - f_1) + g_1 + g_0$$

Для вычисления α воспользуемся тождеством: $b^2 - 3ac = (-b - c)^2 - c(3a + 2b + c)$. Из приведенных выше условий на концах отрезка, используя найденное a, получаем: $(-b - c) = d - f_1 + a = 3(f_0 - f_1) + g_1 + g_0$. Положим $z = -b - c$, $w = \sqrt{b^2 - 3ac}$. Учитывая приведенные соображения, находим $w = \sqrt{z^2 - g_1 g_0}$. Определяя теперь знаменатель в выражении для α из условия для градиента на правом конце отрезка:

$$3a = -2b + g_1 - g_0$$

и выражая b через z:

$$-b = z + c$$

находим следующую последовательность действий для вычисления α

$$z = 3(f_0 - f_1) + g_1 + g_0 \qquad w = \sqrt{z^2 - g_1 g_0} \qquad \alpha = \frac{z + w + g_0}{2z + g_1 + g_0}$$

(предполагалось, что $z^2 - g_1 g_0 > 0$).

В выражении для α знаменатель, равный $3a$, обращается в нуль, т. е. $2z + g_1 + g_0 = 0$ при снижении порядка интерполяционного полинома (при $a = 0$). Преобразуем последнюю формулу для α к виду, пригодному для вычислений при $a = 0$. Запишем выражение $w = \sqrt{z^2 - g_1 g_0}$ в виде $(z + w)(z - w) = g_1 g_0$. Далее последовательно находим:

$$\frac{z+w}{g_0} = \frac{g_1}{z-w} \qquad \frac{z+w+g_0}{g_0} = \frac{z-w+g_1}{z-w}$$

$$\frac{z+w+g_0}{z-w+g_1} = \frac{g_0}{z-w} \qquad \frac{2z+g_1+g_0}{z-w+g_1} = \frac{z-w+g_0}{z-w}$$

Поделив второе соотношение на четвертое, найдем:

$$\alpha = \frac{z+w+g_0}{2z+g_1+g_0} = \frac{g_0}{z-w+g_0}$$

Умножим числитель и знаменатель последней дроби на $(z + w - g_0)$ и воспользуемся тождеством

$$(p - q + r)(p + q - r) = p^2 - (q - r)^2 = p^2 - q^2 - r^2 + 2qr$$

После выполнения преобразований находим

$$\alpha = \frac{z+w-g_0}{2w+g_1-g_0}$$

Убедимся, что это выражение имеет смысл при $a = 0$, т. е. при снижении порядка интерполяционного многочлена (параболическая интерполяция). В этом случае, как мы видели, $2z + g_1 + g_0 = 0$ и из $g_1 = 3a + 2b + c$ следует, что $2b = g_1 - g_0$. Интерполяционный многочлен имеет минимум лишь при $b > 0$, т. е. при $g_1 - g_0 > 0$, а в этом случае $2w + g_1 - g_0 > 0$ и выражение для α имеет смысл. Когда $2w + g_1 - g_0 = 0$, то $g_1 \leqslant g_0$ и, если $a \neq 0$, то применима первая из найденных формул для α:

$$\alpha = \frac{z+w+g_0}{2z+g_1+g_0}$$

а при $a = 0$ из $g_1 \leqslant g_0$ следует, что $b \leqslant 0$, значит интерполяционный многочлен, построенный на основе имеющейся информации, минимума не имеет. В этом случае, а также при

$$z^2 - g_1 g_0 \leqslant 0 \ (b^2 - 3ac \leqslant 0) \qquad \alpha < 0,\ \alpha > 1$$

целесообразно несколько увеличить шаг в данном направлении или (при $g_1 > g_0$) прибегнуть к линейной экстраполяции по градиенту с ограничением величины шага.

Если интерполяция выполняется на отрезке $[0, \gamma]$, $\gamma > 1$ $(\gamma < 1)$, то, произведя сжатие по осям x и y в γ раз (растяжение в $1/\gamma$ раз), перейдем к интерполяции на отрезке $[0, 1]$. При этом преобразовании f_0, f_γ уменьшаются в γ раз (увеличиваются в $1/\gamma$ раз), а g_0, g_γ остаются неизменными. Выполняя затем обратное преобразование, получаем следующие формулы кубической интерполяции на отрезке $[0, \gamma]$:

$$z = \frac{3(f_0 - f_\gamma)}{\gamma} + g_\gamma + g_0 \qquad w = \sqrt{z^2 - g_\gamma g_0} \qquad \frac{\alpha}{\gamma} = 1 - \frac{w - z + g_\gamma}{2w + g_\gamma - g_0}$$

В заключение следует отметить, что во избежание потери точности при расчетах на ЭВМ, вычисление w целесообразно выполнять по формуле

$$w = m \sqrt{\left(\frac{z}{m}\right)^2 - \left(\frac{g_\gamma}{m}\right)\left(\frac{g_0}{m}\right)} \qquad m = \max\{|z|,\ |g_\gamma|,\ |g_0|\}$$

ПРИЛОЖЕНИЕ 3. СИНТЕЗ СИСТЕМ РЕКТИФИКАЦИИ С РЕКУПЕРАЦИЕЙ ТЕПЛА

Имеется ряд работ, посвященных этой проблеме [166]; здесь поставим более ограниченную задачу. Рассмотрим вначале задачу синтеза систем разделения (СР) без

связанных тепловых потоков (на примере четырехкомпонентной смеси, см. рис. 36); при этом будем исходить из следующих предположений: 1) ректификационные колонны (РК) обладают высокой разделительной способностью; 2) давление во всех колоннах постоянно; 3) потоки могут обмениваться теплом только один раз. В дальнейшем для простоты изложения РК, соответствующую вершине A_i^j, также обозначим через A_i^j. Поставим в соответствие вершинам первого типа двоичные переменные z_{ji}, определяемые следующим образом:

$$z_{ji} = \begin{cases} 1, \text{ если схема содержит РК } A_i^j \\ 0, \text{ в противном случае} \end{cases}$$

Не все из этих переменных являются независимыми. Действительно, если существует колонна A_1^3, то существует и колонна A_2^4 и т. д. Отсюда вытекают следующие соотношения

$$z_{31} = z_{42} \qquad z_{32} = z_{43} \qquad z_{12} = z_{23} = z_{24} \qquad z_{37} = z_{46} \qquad z_{38} = z_{47} \qquad (\text{П}, 1)$$

Поскольку в СР может быть только одна из колонн A_1^1, A_2^1, A_3^1, выполняется соотношение

$$z_{11} + z_{12} + z_{13} = 1 \qquad (\text{П}, 2)$$

Колонны A_1^3, A_2^3 и A_7^3, A_8^3 могут присутствовать в СР в том и только в том случае, когда в ней имеются РК A_1^1, A_3^1, соответственно. Отсюда справедливы соотношения

$$z_{11} = z_{31} + z_{32} \qquad z_{13} = z_{37} + z_{38} \qquad (\text{П}, 3)$$

Обозначим через p_{ij} минимальные затраты на РК. Как указывалось выше, эти величины могут быть подсчитаны отдельно при независимой оптимизации каждой колонны. Легко видеть, что задача оптимального синтеза СР формулируется следующим образом: найти минимум функции

$$F^{(2)} = \sum_{i,\,j} p_{ij} z_{ij} \qquad (\text{П}, 4)$$

При наличии ограничений (П, 1)—(П, 3) суммирование в формуле (П, 4) проводится по всем вершинам 1-го типа. Подставляя соотношения (П, 1), (П, 3) в функцию (П, 2), (П, 4) придем к следующей задаче:

$$\min_{z_{ij}} [(p_{11} + p_{31} + p_{42}) z_{31} + (p_{11} + p_{32} + p_{43}) z_{32} + (p_{12} + p_{23} + p_{24}) z_{12} + (p_{13} + p_{37} + p_{46}) z_{37} + (p_{13} + p_{38} + p_{47}) z_{38}] \qquad (\text{П}, 5)$$
$$z_{31} + z_{32} + z_{12} + z_{37} + z_{38} = 1$$

Решение этой задачи сводится к полному перебору, который может быть сокращен применением метода отсечения неперспективных вариантов (см. гл. VI).

Рассмотрим теперь задачу синтеза СР с использованием связанных тепловых потоков. Пусть, как и в предыдущем случае, найдены оптимальные режимы для всех РК всех возможных СР данной смеси. После этого для каждой РК будут известны ее стоимость, а также тепловые нагрузки на холодильник и нагреватель. Обозначим через p_{ij} стоимость РК без учета стоимостей холодильника и нагревателя. Сформулируем подзадачу синтеза теплообменной системы (ТС) в данном случае. Включим в множество холодных потоков S_c возвращаемые части кубового продукта, а в множество S_h — потоки из верха колонн, соответствующих всем вершинам 1-го типа. Так, выходные потоки a и bcd колонны A_1^1 попадут в совокупности S_h и S_c, соответственно. Если поток является выходным, т. е. отбирается из верха (низа) колонн, на вход которых подается одна и та же смесь, то в соответствующее множество его заносят только один раз. Так, поток из верха колонны, содержащий одну компоненту b, присутствует на выходах колонн A_1^3, A_6^4, в которые подается одна и та же смесь bc. В совокупность S_h вносим только один из этих потоков, однако поток, содержащий одну компоненту b, выходящий из верха колонны A_1^3, вводим в совокупность S_h отдельно,

Введем двоичные переменные x_{ij} в подзадачу синтеза ТС [см. соотиошение (VI, 62)]. Построим $(n \times n)$-матрицу назначений для совокупности потоков S_h и S_c, используя известные тепловые нагрузки на нагреватели и холодильники.

Принципиальное отличие данной задачи синтеза ТС от рассмотренной выше состоит в следующем. В обычной задаче синтеза ТС все потоки, входящие в матрицу назначений, обязательно участвуют в теплообмене. В данном же случае заранее не известно, какая совокупность горячих и холодных потоков будет фактически участвовать в теплообмене, поскольку заранее не известна структура СР. С другой стороны структура СР не может быть найдена без знания структуры ТС, стоимость которой составляет существенную часть стоимости всей СР. Например, потоки *bcd* и *a* могут быть использованы в подзадаче синтеза ТС, в том и только в том случае, когда в СР имеется РК A_1^1. Таким образом, необходимо найти условия, связывающие использование того или иного потока в подзадаче синтеза ТС с наличием или отсутствием этого потока в задаче синтеза СР.

Рассмотрим некоторый поток с номером l из совокупности S_h. Пусть этот поток является выходным, т. е. отбирается из верха m РК $A_{q_1}^{p_1}, A_{q_2}^{p_2}, \ldots, A_{q_m}^{p_m}$. Тогда этот поток может участвовать в теплообмене только в том случае, когда в СР присутствует одна из РК $A_{q_i}^{p_i}$, $(i = \overline{1, m})$. Это условие может быть записано следующим образом:

$$\sum_{k=1}^{N} x_{lk} = \sum_{i=1}^{m} z_{p_i q_i} \qquad (\text{П}, 6)$$

Легко проверить (см. рис. 35), что в одной СР не может быть двух РК, которые бы давали один и тот же поток, поэтому из совокупности переменных $z_{p_i q_i} (i = \overline{1, m})$ только одна может стать равной единице, а остальные будут равны нулю. Следовательно, правая часть условия (П, 6) может быть равна только нулю или единице. Таким образом, если в СР присутствует одна из РК $A_{q_i}^{p_i} (i = \overline{1, m})$, то правая часть условия (П, 6) будет равна единице, откуда следует, что одна из переменных x_{lk}, $(k = \overline{1, N})$ равна единице, т. е. теплообмен потока l разрешен. Если в СР нет ни одной РК из совокупности $A_{q_i}^{p_i} (i = \overline{1, m})$, то правая часть условия (П, 6) будет равна нулю и теплообмен этого потока с другими будет отсутствовать.

Рассмотрим пример. Пусть поток с номером l из S_h содержит одну компоненту b, которую получают разделением потока *bc*. Как поток с верха РК, он может появиться на выходе РК A_1^3, A_6^4. Соотношения (П, 6) в данном случае примут вид

$$\sum_{k=1}^{N} x_{lk} = z_{31} + z_{46}$$

Соотношения (П, 6) должны быть выписаны для каждого горячего и холодного потока. Теперь задачу синтеза СР со связанными тепловыми потоками можно записать следующим образом:

$$\min_{x_{ij},\ z_{ij}} \left(F^{(1)} + F^{(2)}\right) \qquad (\text{П}, 7)$$

где $F^{(1)}$, $F^{(2)}$ определяются по формулам (VI,6), (П, 4), соответственно, а переменные z_{ij}, x_{ij} удовлетворяют соотношениям (П, 5) и всем соотношениям типа (П, 6).

Заметим, что избавиться от допущения 3 можно так же, как и в обычной задаче синтеза ТС. Отказ от допущения 2 приведет к итерационной двухуровневой процедуре, первому уровню которой будет соответствовать решение задачи (П, 7), а второму — изменение давлений в каждой РК. Отметим, что приведенная процедура не дает полного решения рассматриваемой задачи, поскольку коэффициенты p_{ij}, f_{ij} определялись на основе данных, не учитывающих взаимного влияния РК через связанные тепловые потоки.

Геннадий Маркович Островский,
Тевель Александрович Бережинский

ОПТИМИЗАЦИЯ ХИМИКО-ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ. ТЕОРИЯ И ПРАКТИКА

Редактор *Р. М. Степанова*
Художник *Н. В. Носов*
Художественный редактор *Н. В. Носов*
Технический редактор *О. В. Тюрина*
Корректор *Н. А. Иванова*

ИБ № 1268

Сдано в набор 10.08.83. Подп. в печ. 06.02.84. Т-06304.
Формат бумаги $60 \times 90^1/_{16}$. Бумага тип. № 3. Гарн. литературная.
Печать высокая. Усл. печ. л. 15,0. Усл. кр.-отт. 15,0.
Уч.-изд. л. 16,63. Тираж 4800 экз. Заказ № 240.
Цена 1 р. 20 к Изд. № 2176

Ордена «Знак Почета» издательство «Химия», 106076, Москва, Стромынка, 13

Ленинградская типография № 6 ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 193144, г. Ленинград, ул. Моисеенко, 10.

